



# TM 12-260

WAR DEPARTMENT TECHNICAL MANUAL

---

## PERSONNEL CLASSIFICATION TESTS

THE ARMY LIBRARY

WASHINGTON, D. C.

THE ARMY LIBRARY

---

WAR DEPARTMENT • APRIL 1946



WAR DEPARTMENT TECHNICAL MANUAL

TM 12-260

*This manual supersedes TM 12-260, 31 December 1942.*

---

# PERSONNEL CLASSIFICATION TESTS

THE ARMY LIBRARY

WASHINGTON, D. C.

---



WAR DEPARTMENT

APRIL 1946

THE ARMY LIBRARY

---

*United States Government Printing Office*

*Washington: 1946*

WAR DEPARTMENT

Washington 25, D. C., 26 April 1946

TM 12-260, Personnel Classification Tests, is published for the information and guidance of all concerned.

[AG 300.7 (5 Mar 46)]

BY ORDER OF THE SECRETARY OF WAR:

OFFICIAL:

DWIGHT D. EISENHOWER  
*Chief of Staff*

EDWARD F. WITSELL  
*Major General*  
*The Adjutant General*

DISTRIBUTION:

AAF (10); AGF (10); ASF (2); T (5); S Div ASF (2); SvC (2); GH (5); SH (2); RH (2); Gen & Sp Sv  
Sch (10); Tng C (15); RC (10); Reclss C (5); Rehab C (2); Sep C (5); Ind Dis (5); A (5); CHQ (5);  
D (5); B (5); R (2); Bn (2); W (5); G (2). T/O & E 8-590 (2); 20-42 (5).  
Refer to FM 21-6 for explanation of distribution formula.

## FOREWORD

---

This manual provides essential information about the development, interpretation, and correct use of personnel classification tests and results. For proper operation of the personnel classification system as a whole at all echelons, knowledge of AR 345-5, TM 12-425, 12-426, 12-427, 12-405, 12-406, and 12-407 is required.

The Adjutant General is the War Department operating agency charged with the development, construction, validation, and standardization of all Army personnel screening tests, except air crew tests. (See Sec. II, WD Circular 312, 1943.) Only tests which have been authorized by The Adjutant General are used in classifying military personnel. Such tests are described in this manual and changes thereto. Authorized tests are also listed in monthly issues of FM 21-6 which prescribes the procedures for requisitioning tests and test supplies, and indicates the distributing agencies from which they may be procured.

Field installations recognizing a need of new tests for special purposes are encouraged to submit their problems through appropriate channels to The Adjutant General, Attention: Classification and Replacement Branch, Personnel Research Section. The use of military personnel to try out new or unauthorized tests is not permitted unless the purposes of the experiment and the procedures to be followed have first been considered by The Adjutant General and received his approval.

This manual has been published in loose-leaf form to facilitate changes. Such changes will be supplied on a page basis, and will be published as required. As change pages are received they will be inserted in their proper places, and the replaced pages destroyed.

Each page of the manual bears the date of publication in its upper inside corner. Pages which represent changes will carry the date and number of the change.

Paragraphs are numbered consecutively throughout the entire manual. Paragraphs carrying decimal suffixes will indicate newly added material; for example, a paragraph numbered 23.1 will represent the first main paragraph following paragraph 23.

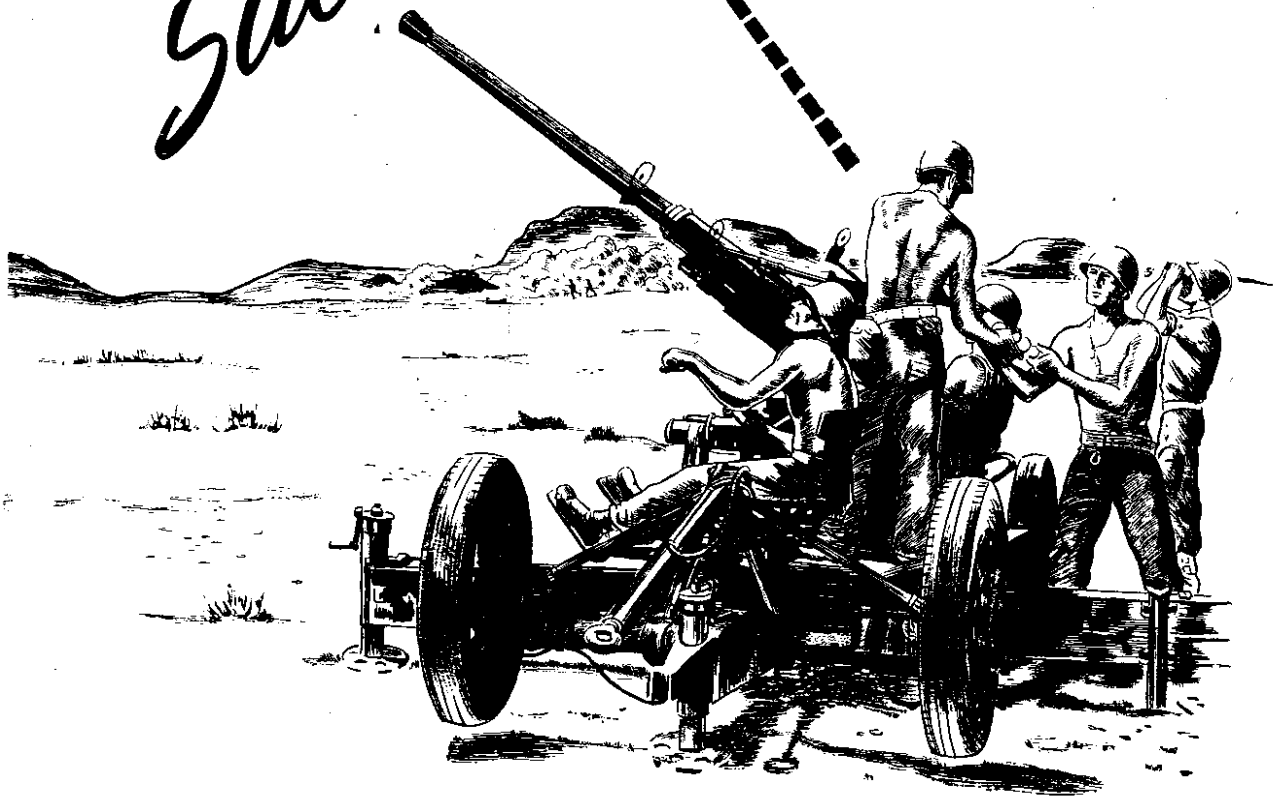
Pages are numbered consecutively throughout the manual. If new pages are added, these will carry alphabetical suffixes. For example, if a new page is added between 51 and 52, this page will be numbered 51-A.





..... THE RIGHT MAN  
AT THE RIGHT PLACE  
AT THE RIGHT TIME

*Success in Battle*



# CONTENTS

	Paragraph	Page
CHAPTER 1. INTRODUCTION.....	1-7	1
CHAPTER 2. ARMY TESTS AS SCIENTIFIC TOOLS.		
Section I. Measurement problem.....	8-9	5
II. Measuring skills and aptitudes.....	10-14	9
III. Army tests as scientific tools for measuring skills and aptitudes.....	15-19	11
CHAPTER 3. HOW ARMY CONSTRUCTS TESTS.		
Section I. Planning test.....	20-26	13
II. Constructing test.....	27-30	17
III. Trying out test.....	31-34	19
IV. Establishing the scale of measurement.....	35-38	22
V. Summary.....	39	24
CHAPTER 4. ADMINISTRATION AND SCORING OF TESTS.		
Section I. Introduction.....	40-42	25
II. Principles and procedures for administering group tests.....	43-45	27
III. Directions for scoring group tests.....	46-51	31
IV. Directions for administering individual tests.....	52-54	36
V. Scoring individual tests.....	55-57	38
VI. Obtaining standard scores.....	58-59	39
CHAPTER 5. MEANING AND INTERPRETATION OF TEST SCORES.		
Section I. Army standard score scale.....	60-64	41
II. Significance of test scores.....	65-69	45
III. Reliability.....	70-72	49
IV. Summary.....	73	50
CHAPTER 6. RATING SCALES, QUESTIONNAIRES, AND INTERVIEWS.		
Section I. Orientation.....	74-79	51
II. Rating scales.....	80-84	53
III. Questionnaires and check lists.....	85-90	56
IV. Interview.....	91-96	59
CHAPTER 7. AUTHORIZED ARMY TESTS — USE IN INDUCTION STATIONS, RECEPTION CENTERS, AND OTHER ARMY INSTALLATIONS.		
Section I. Psychological testing in induction station.....	97-107	62
II. Psychological testing in reception center.....	108-113	68
III. Psychological testing in other installations.....	114	75
CHAPTER 8. AUTHORIZED ARMY TESTS — USE IN SELECTION AND ASSIGNMENT.		
Section I. Orientation.....	115-120	76
II. Administrative and clerical assignments.....	121-124	78
III. Automotive and mechanical assignments.....	125-132	79
IV. Miscellaneous technical specialties.....	133-138	84
V. Other tests.....	139-140	86
INDEX.....		88



*This manual supersedes TM 12-260, 31 December 1942*

## CHAPTER I

### INTRODUCTION

#### 1. Classification and Modern War

Modern warfare has outmoded the old-fashioned all-purpose soldier. The Army of today is a "team of teams" each of which is made up of men who depend on one another to do their particular and often specialized jobs. Inefficiency in combat is paid for in human lives. It is, therefore, of vital importance to place in every assignment men who are physically, emotionally, and mentally qualified to do what is required. Every officer who is responsible for deploying troops or for making direct man-from-man assignments should be able to find out which soldiers are capable of doing each of the many Army jobs, and how well they can be expected to do them. Whether or not officers are able to learn these things readily and accurately depends mainly on two factors:

- a. The efficiency of the Army's system of personnel classification.
- b. The officer's own understanding of classification information, and his ability to apply it to his practical problems of utilization of men.

#### 2. Purpose of Classification

The millions of men and women who make up the Army possess different combinations of skills and abilities. They represent experience in some eight thousand civilian occupations. There are men who can build a bridge or drive a nail, men who can plow a furrow, or plead a case. There are men who can plan and organize, and there are men who have skill and strength in their hands. All of them must be apportioned among more than five hundred types of Army jobs, each of which requires a somewhat different combination of skills, aptitudes, and training. The task of the classification system is to discover the military abilities of these men and their aptitudes—or "trainability"—for Army jobs. This information must then be analyzed, recorded and passed on, with systematic recommendations concerning assignment.

#### 3. Scope of This Manual

Personnel research, the subject of this manual, is the scientific work which makes sound classification

possible. Personnel research has the following main missions:

- a. To develop and evaluate tests, rating scales, and interview methods which are the instruments used for finding out how much of the various skills and aptitudes required for Army jobs and Army training each soldier may possess.
- b. To develop methods for expressing and recording each man's skill and aptitude in the most useful way possible.
- c. To inventory the abilities found among soldiers as an aid to selection, training, deployment and redeployment.

#### 4. Need for Scientific Methods

The proper deployment of abilities is as essential to success in battle as the tactical disposition of the men and matériel. Every efficient division represents a balanced assembly of skills, aptitudes, and physical characteristics, each present in sufficient numbers for the task at hand. As attrition and casualties thin the ranks, these skills and aptitudes must be replaced at once and in kind. There must be an unceasing flow across enormous reaches of terrain and water, and this replacement stream must always contain the various kinds of men that may be required at a particular place and time. Both the building of units and their constant repair by efficient replacement depend upon precise knowledge concerning millions of men. Only scientific methods can furnish such knowledge. Personal judgments take too much time, and vary too much among the persons making them. They are too difficult to record and pass along in a form that can be readily and correctly interpreted. The scientific methods of personnel research enable officers to know their men on short acquaintance. They enable higher commanders to know in detail the forces which are created in a few brief months out of an unsorted mass of peaceful civilians and replenished from a "manpower barrel" which is by no means bottomless.

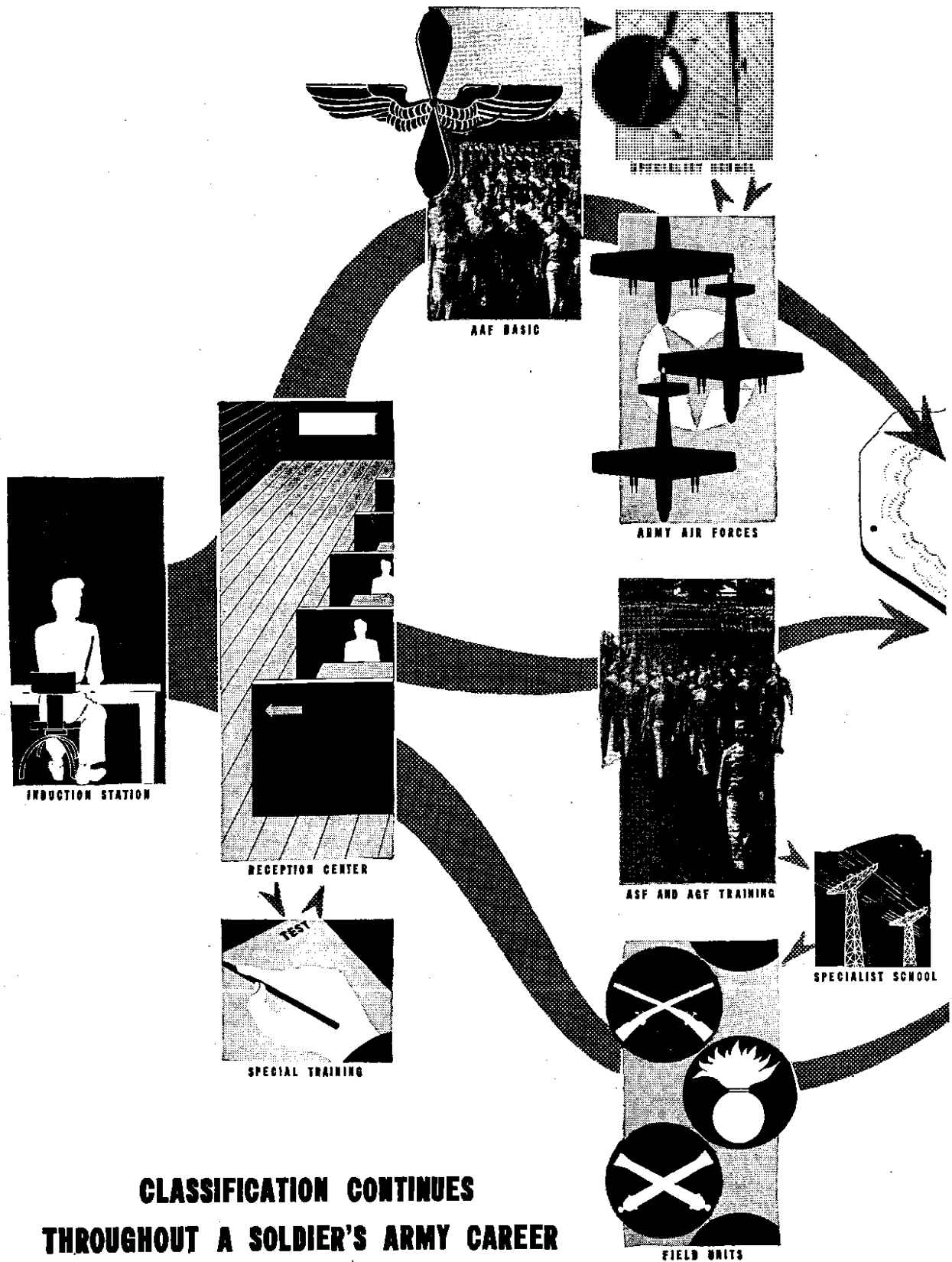


Figure 1.

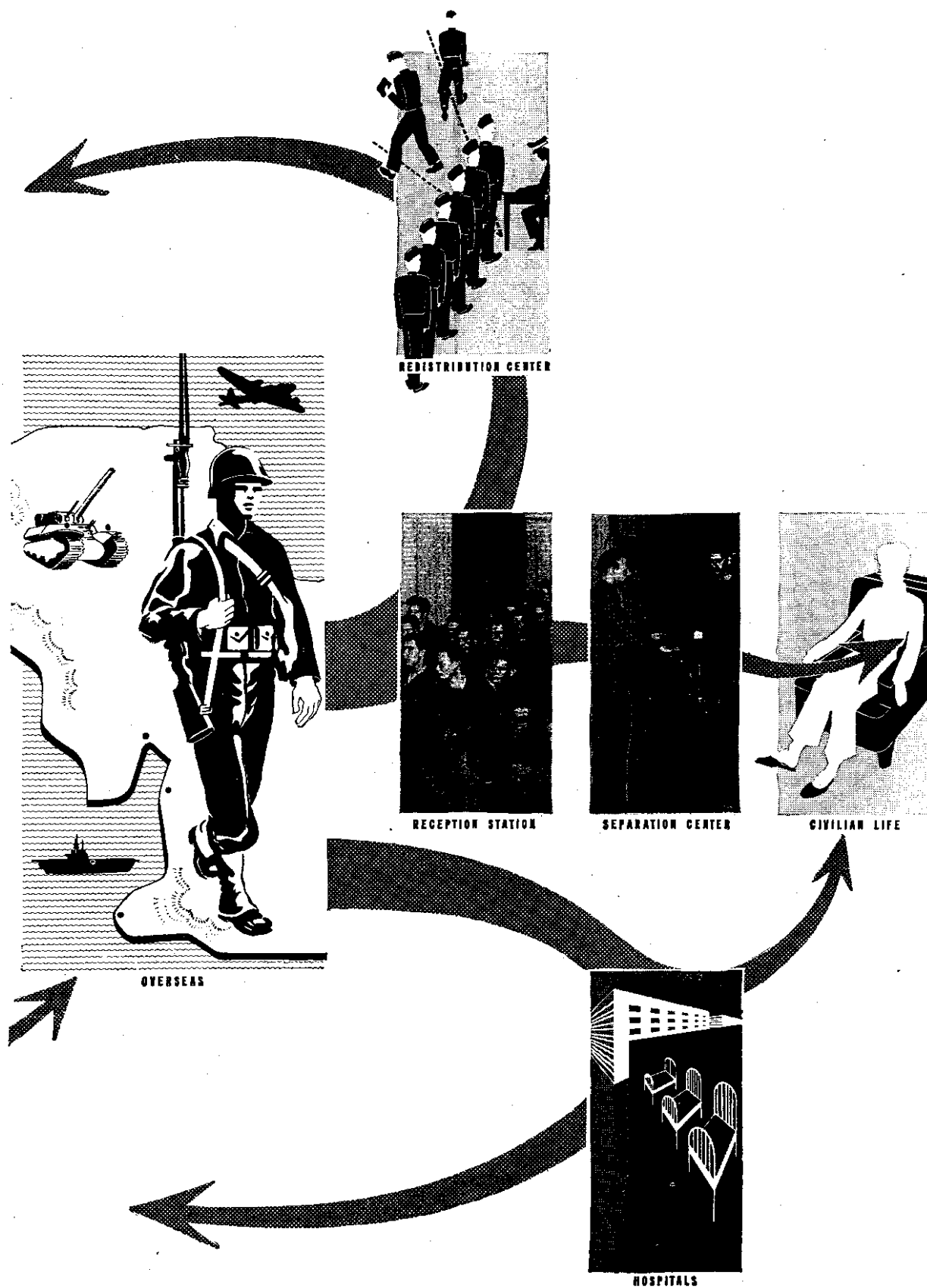


Figure 1—Continued.

## 5. Ability is an Equation

It happens seldom, if ever, that one characteristic is the key to success in performing an Army assignment. A man with a "strong" mind and a weak back will not be fit for duty where he will have to dig fox holes; a man with both a "strong" mind and a strong back will not be satisfactory in combat if he faints at the sight of blood. Ability is an equation in which all important characteristics are balanced against the job. On one side of the equation is the assignment itself, broken down into the demands it makes on the individual. On the other side are the characteristics of the men being considered. These characteristics are of the following general kinds:

a. *Physical characteristics*, such as strength, endurance, agility, defects, malformations, and other bodily traits which have a bearing on ability to perform assignments.

b. *Emotional characteristics*, such as tendencies toward depression, split personality, and other weaknesses which may make a man crack up under the pressures of training or the rigors of combat.

c. *Mental characteristics*, such as ability to learn, and skill acquired through education, training, and experience.

## 6. Measure of a Man

The three kinds of characteristics listed in paragraph 5 must all be measured or reliably estimated before the Army knows whether a man is soldier material. It is necessary to measure some (and sometimes all) of them at various stages of a soldier's service to determine what special training it is profitable to give him and what assignments he can be expected to perform satisfactorily. Throughout his military career, from basic training to combat and return (fig. 1), his abilities must be continually reviewed and evaluated. As he acquires new

military skills through training and experience, or as the changing needs of the Army demand, his assignment is subject to revision. Each change is made in the light of fresh information about the soldier and up-to-date requirements of the tactical situation.

## 7. Usefulness of This Manual

Personnel research provides the procedures (tests, rating scales, etc.) used by the Army to measure mental characteristics. These procedures are developed to meet practical needs and designed for use under those conditions usually met with in Army installations. They are in most cases special-purpose instruments. That is, they must be used to obtain only the information they were designed to discover, and they must be used in the proper fashion. Moreover, the results (the scores, ratings, etc., which describe the abilities of men and predict how well they will do Army jobs) must be interpreted properly if they are to have any practical value. This manual explains how measurement procedures are developed in the Army, how they are to be used, and how to apply the results to practical problems of selection and assignment. It is directed mainly to those officers and men who are principally concerned with the technical phases of the classification system—the classification officers, personnel consultants, clinical psychologists, and the enlisted classification specialists in all echelons. But it will also aid all officers who are responsible for the conditions under which the work of classification is carried on and its products utilized. Officers responsible for the disposition and deployment of skills and abilities of men, and those who make assignments, will find here a key to one of the Army's most valuable assets—the classification data concerning every soldier in the United States Army.

## CHAPTER 2

### ARMY TESTS AS SCIENTIFIC TOOLS

#### Section I. MEASUREMENT PROBLEM

##### 8. General

Classification begins with jobs. The five hundred-odd Army assignments, each created by practical necessity, determine the kinds of tests to be built and the use to be made of their findings. Once a definable job is recognized as necessary by the War Department, it is added to the table of organization of the unit in which the specialty is required. Assignments listed in tables of organization are analyzed to define clearly and in detail just what a man is expected to do, and the resulting job descriptions are added to TM 12-427. The preparation and revision of these job descriptions go on continuously as new assignments are developed to meet the changing needs of the Army. The furious rate of growth during recent years is well illustrated by the fact that 22 new military jobs have been added since 1940 to enable the Army to maintain efficient radar installations. The degree of proficiency required in each assignment must be ascertained and training courses established when required to enable men to meet job requirements. Classification is the process of selecting for each assignment the men most likely to reach required proficiency. There are three general types of assignment, each of them requiring different techniques applied during the process of classification and selection. (See ch. 7, 8, and 9 for specific techniques and points at which they are applied.) The types of assignments are as follows:

a. Assignments for which almost any inductable man can reach required proficiency after a short period of practice.

b. Assignments, such as truck driver and automobile mechanic, where skill acquired as a civilian may be sufficient after brief indoctrination and on-the-job training.

c. Assignments for which weeks or months of intensive and special training are required to bring men up to requisite efficiency. The range of these training courses and their importance to the modern Army is illustrated by the facts in table I.

##### 9. Individual Differences

The men who must be selected and assigned to all the training courses and Army occupations are

a varied lot. Even after screening by the physical examination at the induction station, soldiers differ widely in health, strength, size, and endurance. The Army contains men who can march 50 miles a day with full field equipment, and men who could scarcely cover a few miles under the same conditions.

*Table I. Number of special training courses for officers and enlisted men offered by various arms and services*

Arm or service	No. of courses for officers	No. of courses for EM
Army Air Forces:		
Technical Training Command.....	21	116
Army Ground Forces:		
Antiaircraft Artillery.....	11	10
Armored Force.....	8	6
Cavalry.....	6	9
Coast Artillery.....	4	9
Field Artillery.....	10	11
Infantry.....	6	6
Parachute School.....	4	4
Tank Destroyer.....	6	6
Army Service Forces:		
Adjutant General's Department.....	3	5
Army Exchange Service.....	1	0
Corps of Chaplains.....	1	0
Chemical Warfare Service.....	13	6
Corps of Engineers.....	11	20
Finance Department.....	4	3
Judge Advocate General's Department.....	1	0
Medical Corps.....	24	19
Ordnance Department.....	25	40
Provost Marshal General.....	5	2
Quartermaster Corps.....	13	12
Signal Corps.....	16	57
Special Services.....	2	1
Transportation Corps.....	2	0
Totals.....	197	342

Soldiers differ in such characteristics as resistance to extremes of temperature, efficiency at high altitudes, and ability to see at night. In regard to this latter characteristic, which is so important in modern warfare, tests under field conditions show a wide range in the ability to recognize objects (tanks, trucks, howitzers, machine guns, etc.) at night. In one such test, a few men could recognize the object (a  $\frac{1}{4}$ -ton truck) at a distance of 90-99

# DISTRIBUTION OF MEN ACCORDING TO THE DISTANCE AT WHICH THEY WERE ABLE TO RECOGNIZE A ¼ TON TRUCK AT NIGHT

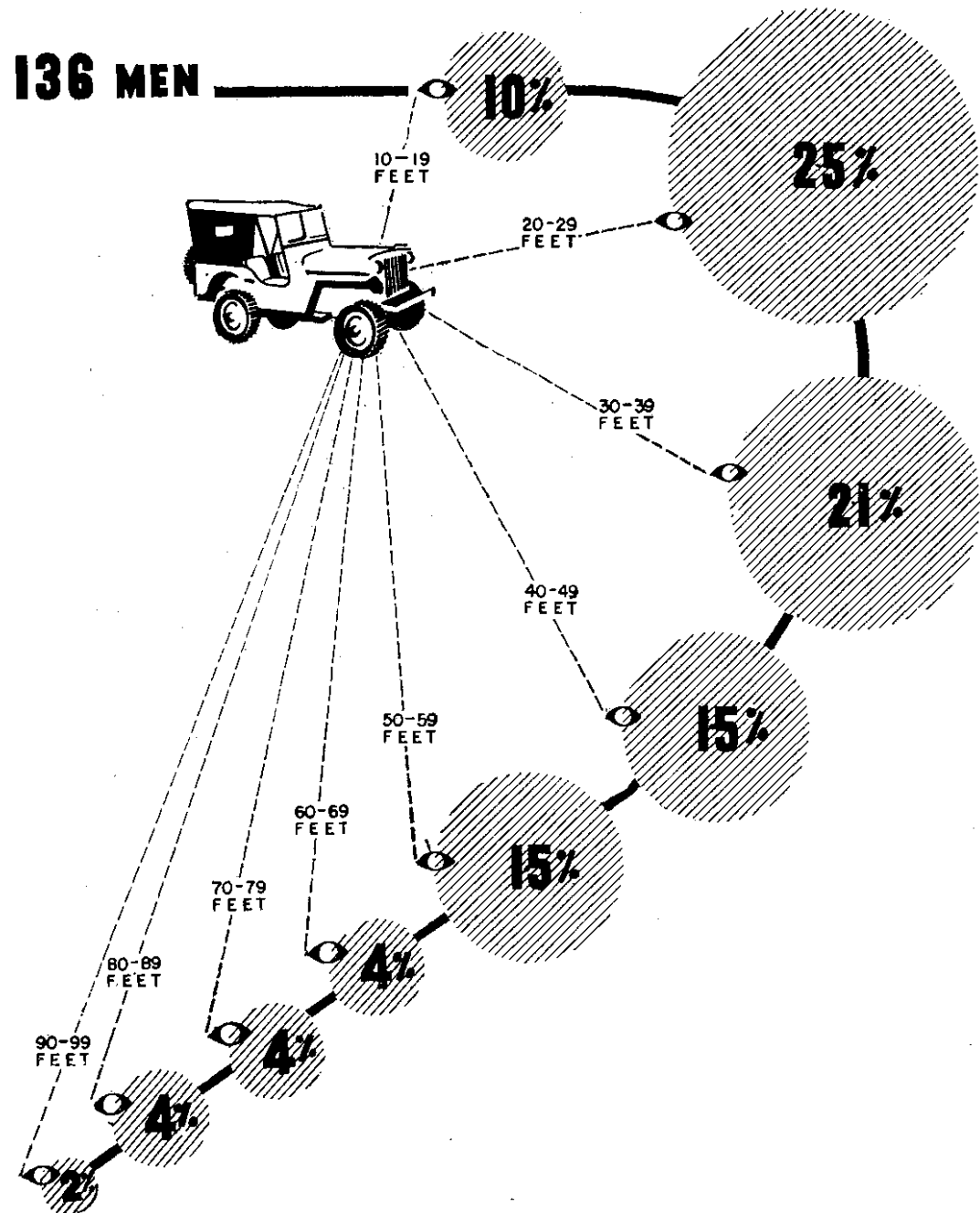


Figure 2.

feet, while others were unable to do so until they were no further than 10 feet away. (See fig. 2.)

a. The differences between soldiers in aptitudes and skills are just as large and as important as their

differences in physique, stamina, and sensory acuity. These psychological characteristics are not so directly observable as physical traits. One cannot tell by looking at a man, or by determining his physical dimensions, whether he can add or spell, drive a nail or repair a carburetor. Nor, what is of more importance, can one tell by such direct observation whether the man can learn to do these things in the relatively short time available for training. Because these abilities and aptitudes are not directly observable they could, without the benefit of scientific techniques, be overlooked in selecting men for specialist training or assigning men to Army jobs. Yet, in every respect, the range of skills and aptitudes among Army men is tremendous. This holds true for general overall ability and for the hundreds of special skills that are required by the Army. Not all men, for example, possess the same capacity for absorbing basic military training. Figure 3 illustrates the distribution of men in one training unit rated according to soldier performance; that is, according to their value to the unit as soldiers. Most of the men are average soldiers. In fact, this is the real meaning of the term "average." Very few men are rated as of no value or of outstanding value; most men fall into the middle—the average—group. Men also tend to fall into definable groups when ranked according to specialized skills. Figure 4 shows the performance of the men in a class for

training radio code operators. After 8 weeks of training, some of the men were receiving code messages at the rate of 18-20 words per minute, while others were not able to do better than a rate of 2 words per minute.

b. Men differ enormously not only in the rate at which they can acquire new skills, but also in the level of skill they can reach with training. Contrary to popular belief, individual differences are not ironed out by training. Even if endless time were available, it is doubtful if the poorest performers could be brought up to the level of the best. On the contrary, practice is more likely to accentuate differences in performance. Those who are more skillful to start with will usually improve faster than the less skillful, with the result that the range of abilities after training is even greater than before. Figure 4 shows that, although the whole class did better after 12 weeks of training than at the end of 8 weeks, the differences in code receiving speed had become even more striking.

c. The fact that each individual possesses more of some skills and aptitudes than of others is equally important to the Army. The Army cannot afford blundering which results in such folly as the waste of a morale-building cook to make a good truck-driver or failure to recognize and train a man whose endowments might have made him a platoon sergeant capable of saving his whole outfit in a critical hour. Such errors must be reduced to a minimum.

#### DISTRIBUTION OF PERFORMANCE RATINGS OF A GROUP OF TRAINEES

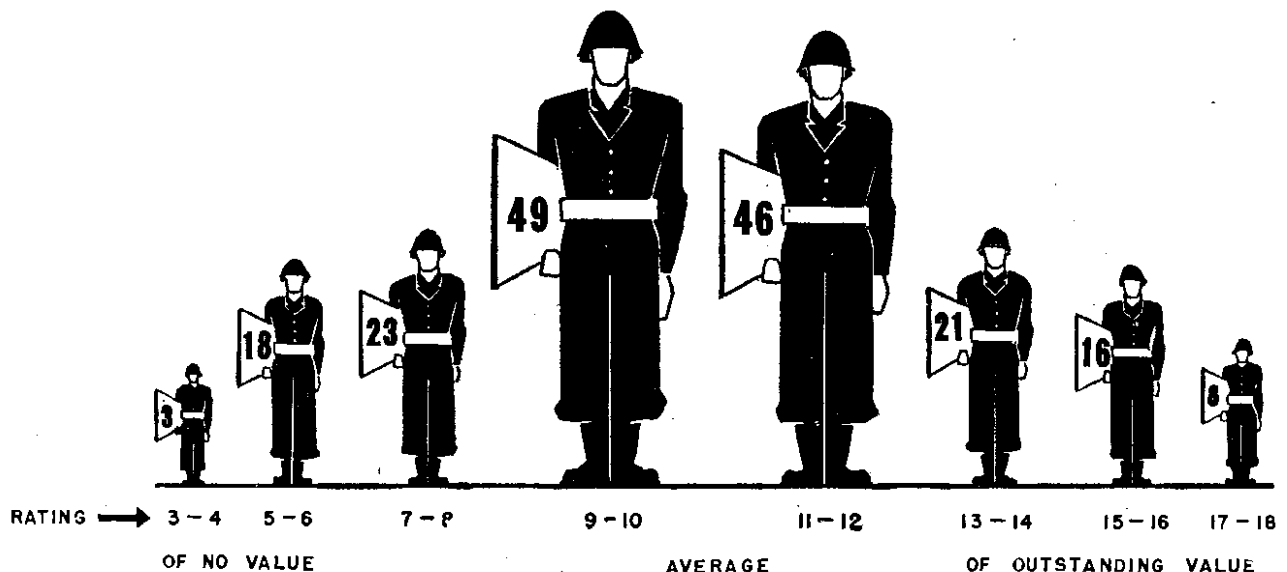


Figure 3.

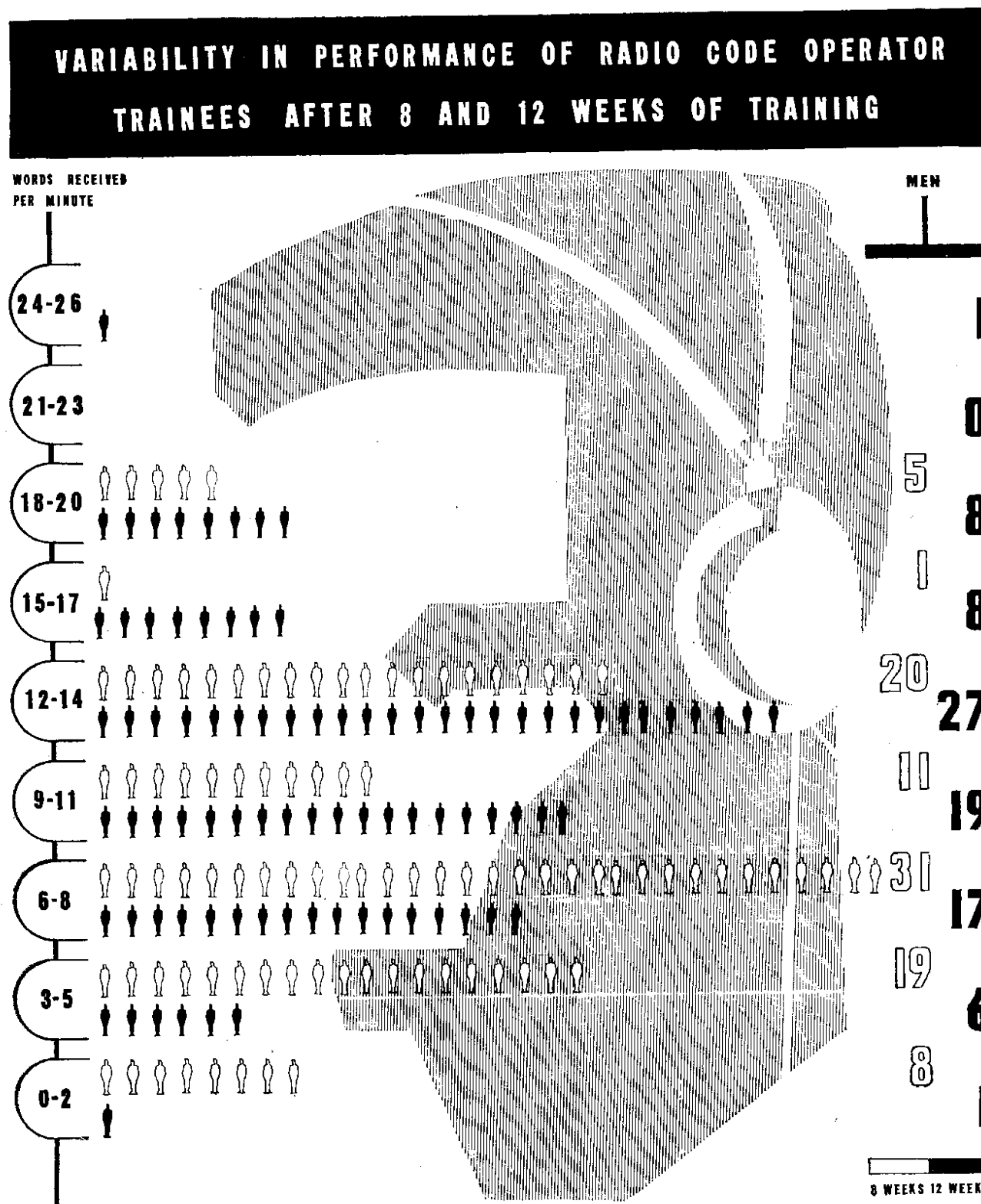


Figure 4.

*The Army must choose for training assignments the men who will respond most readily to that particular*

*training. These principles are self-evident. But the methods successful in selection are anything*



but obvious, rule-of-thumb methods. They were not, in fact, available until very recently. For, to select the best, all men must be measured; and reliable, accurate, and useful techniques for measuring the varying, and often hidden, traits of men are almost as modern as radar. The fact that these techniques and the effective methods of applying them are so new, heightens the importance of a full understanding of principles and procedures on the part of all responsible personnel.

## Section II. MEASURING SKILLS AND APTITUDES

### 10. General

The problem of determining the skills and aptitudes of men is not new. It has been encountered wherever some men have been forced to make judgments about others. Yet it has not always been as urgent as it is now. In other circumstances, when a man's ability to perform a job is questioned, he can be tried on that job. If he succeeds, he must have the necessary ability; if he fails, he can be tried out on a different job. Such trial and error methods, though straightforward, are feasible only when the importance of getting the most competent men on the job far outweighs the tremendous waste in time and effort. They are not practicable, as a general rule, in classification work in the Army. With millions of men to be classified, selected, and assigned, and with training time and facilities at a premium, it is obviously impossible to try each man on the hundreds of Army jobs in order to discover the one to which he should be assigned. Techniques are required by means of which the abilities of men in large numbers can be determined in advance of assignment. And it is also essential that these findings be dependable, yet obtained with a minimum expenditure of time.

### 11. Traditional Methods

Because the psychological characteristics of men are not directly observable, it was inevitable that attempts should be made to find other signs of these characteristics. Through the course of time, numerous systems of "character analysis" have evolved. None of them was systematically developed nor consciously evaluated. Rather, each simply grew as a part of the folklore, and each attracted its own circle of devotees. The earliest of such systems were cloaked in mystery and ritualistic hocus-pocus. They looked not to the

man himself, but read his character and destiny in some outside circumstance—the constellations, the entrails of birds, or in the hallucinated visions of a crystal gazer. These cults are, of course, beyond the realms of reason. In more recent times, however, a series of pseudo-scientific methods have gained some popular appeal. These are usually based on the assumption that psychological characteristics can be detected through various physical signs or symptoms.

a. The most widely held of all these methods is that based on the assumption that the face is an index of the inner man—that traits and skills are reflected in the size and arrangement of the features. Thus, a high forehead is supposed to indicate intelligence, and red hair, a "fiery" temperament. Most of these beliefs are obviously based on far-fetched generalizations or analogies ("men with large ears are inquisitive" and "a fox-like face indicates trickiness"), and are seen to be patently absurd when stated in formal terms. Nevertheless, it is a temptation to "size up" individuals as bright or dull, crafty, sensitive, or humorous according to the shape or size of their facial features. When these haphazard judgments are used to determine assignment or disposition, the situation may become serious. Accumulated evidence, obtained under carefully controlled scientific conditions, has been able to lend no grain of truth to these simple beliefs.

b. A closely allied system of beliefs is that known as phrenology, which claims that a man's make up—his skills and talents, personality and character, can be appraised by judging the relative size and position of irregularities of his cranium. Unfortunately for the system, the brain is not a mosaic of traits; it does not grow with exercise, like a muscle, and its size and shape do not determine the size and contours of the external surface of the skull.

c. Graphology, which purports to read the nature of men in their handwriting, has a certain superficial plausibility. It claims (to choose typical examples) that the clear thinker writes clearly, the forceful personality with a bold stroke, and the ambitious individual on a line slanting upward. As in most of these "character analysis" systems, this is nothing but analogy—a seductive but dangerous way of reasoning. Actually, no one has ever been able to demonstrate the validity of graphology in a properly controlled experiment. In several such experiments, "expert" grapholo-

gists achieved no better than chance success in matching samples of handwriting to the persons who wrote them.

## 12. Indirect Observation

Modern science looks to the men themselves, rather than to extraneous characteristics. As has already been pointed out, it is not practicable to observe the men directly on the job in order to discover whether or not they possess the requisite abilities. They must be evaluated by the method employed throughout all science; that is, by indirect observation which makes it possible to predict performance. One method employed by the Army is to evaluate the previous occupational and educational history of men, since what one has done is a clue to what may be expected of him. Interviews and personal history questionnaires therefore have a part in classification and assignment procedure. (See ch. 6.) They are seriously limited by the fact that they are time consuming and lacking in complete objectivity. The examiner may fool himself and the examinee may fool him. Moreover, questionnaires and interviews lack precision, not only because military assignments differ from civilian occupations, but also because they cannot result in dependable comparisons between man and man nor provide the exact data required for prediction. It is necessary to sample the performance of soldiers by tests which measure and predict in a truly scientific fashion.

## 13. Scientific Measurement

Progress in science goes hand in hand with the development of measuring instruments. Primitive man could weigh an object by "hefting" it to judge how heavy it seemed; the weighing of bits of matter too small to be seen awaited the development of incredibly delicate instruments. The early physician judged the temperature of his patient by placing a hand on his fevered brow. Today he puts a thermometer in the patient's mouth and notes the height of the column of mercury along a scale. Modern measurement is taken for granted, but it has several important characteristics that must be kept clearly in mind before classification testing can be understood.

a. The measurement is indirect. The phenomenon or characteristic being determined is measured by noting its effect on some other phenomenon that can be observed. Temperature, for example, cannot be seen, nor can it be directly sensed with

any degree of accuracy. Yet because the physicist has established a constant and invariable relation between temperature and the expansion of mercury, the simple and direct observation of the thermometer can be used as a measure of temperature. Moreover, this measure of temperature is itself obtained because it is an indirect indication of something else that the physician is really concerned with, namely, the patient's health.

b. The measurement is objective. That is, the result obtained is almost completely independent of the person doing the measurement. Experience shows that subjective estimates and judgments are influenced by many factors that have to do with the observer rather than the phenomenon he is observing. An object may feel warm or cool depending upon the observer's own temperature or upon his expectations, desires, suggestibility, prejudices, or a number of other irrelevant factors. The thermometer has no personal bias. It will yield the same result to all observers providing only that they can read it properly.

c. The measurement is reliable. It does not produce one result at one time and something quite different a moment later. This consistency is necessary in measurement, since without it there is no way of knowing whether a change in results may be confidently attributed to a real change in the phenomenon measured or to a variation in the instrument.

d. The measurement is sensitive. It permits the discovery of small variations or fine discriminations in the characteristic being measured. Crude, unaided judgment can easily distinguish between extremes of hot and cold. But, few men could tell the difference between temperatures of 99° and 102° without the aid of a thermometer; and it is precisely such small variations that are of critical importance in medicine.

e. The measurement is *meaningful*, in the sense that it can be interpreted correctly and usefully by any trained person. Any given reading on a good instrument always means the same thing because it is higher or lower or more or less than a standard reference or "bench mark" on a scale which is also standardized. On the centigrade thermometer, for example, any obtained reading can be interpreted by reference to the temperature at which water freezes, and the temperature at which it boils. This is possible because the thermometer is scaled to these two "reference points" which are definite and independent of the measuring

instrument. In contrast, such subjective estimates as "fairly cool" or "very hot" are merely personal opinions which may mean quite different things to different people and are therefore of little use in measurement.

#### 14. Army Tests as Scientific Tools

Scientific measurement is applied to the problem of determining the skills and aptitudes of men in the Army. Though psychologists use paper-and-pencil tests for the most part rather than mechanical devices, the tests are nevertheless accurate and useful instruments analogous to thermometers, scales, and other aids to physical measurement. Those developed and used by the Army in connection with the selection and assignment of men yield results that are objective, reliable, and meaningful. Army tests are especially constructed to serve as the measuring instruments of the classification system. As such, they are used to determine the skills and aptitudes of Army men in order that they can be selected for training courses and assignments that are suited to their abilities. These tests are used because experience has amply demonstrated that they result in better selection; they save the Army time, money, and facilities and make all echelons more efficient. The reasons for their superiority over other techniques are to be found in the fact that the Army tests have the same characteristics as all scientific measuring devices.

### Section III. ARMY TESTS AS SCIENTIFIC TOOLS FOR MEASURING SKILLS AND APTITUDES

#### 15. Indirect Measurement of Traits

As this chapter has already shown, psychological traits cannot be observed directly. But the products of these traits can be reviewed, examined, and evaluated. In other words, the differences between men in the amounts of some skill or capacity they possess can be determined by measuring and evaluating their differential performance on tests involving those skills. A soldier's skill in mechanics can be determined by observing his performance on mechanical tasks or his response to questions involving mechanical principles and practices. The test poses questions and problems which, for their correct solution, involve the characteristics or traits to be measured. The number of such questions and problems which the soldier

can respond to correctly is an indirect measure of the amount of the trait he possesses.

#### 16. Uniformity

The questions that may be asked about a topic are almost unlimited in number, and can be phrased in many different ways. It follows that questions which are not uniform lead to answers so various that they reveal nothing useful in comparing men with one another. It is characteristic of informal questioning techniques that now some questions are asked, now others; that the manner of wording or presenting the questions varies; and that the interviewer is at one time tolerant and at another time critical in his evaluation of answers. The test, on the other hand, is composed of a standard series of questions selected by careful scientific techniques. (See ch. 3.) Thus, all individuals to be tested are given identical questions or tasks, presented in a uniform prescribed manner. Moreover, in the typical Army test, the method of indicating answers is simplified to the point where they can be scored by mere counting or by a wholly impersonal machine. The whole process is analogous to that of applying a scale and reading off the result. Any reasonably experienced examiner can obtain the same measurement by following instructions.

#### 17. Reliability

No measuring instrument will produce results with perfect consistency. Even a series of successive measurements of a wooden plank will contain variations. But for satisfactory carpentry, any one of these estimates is close enough to be considered the "true" length. Should any two measures differ by as much as a few inches, however, the carpenter's confidence in his rule is such that he would sooner conclude that he was measuring different planks than he would suspect the constancy of his rule. Psychological measurement is perhaps not so consistent as this; the plank is not so fickle as the men. Yet the estimates of the talents of men obtained with the use of Army tests are sufficiently close to their "true" measures to be employed with confidence for the classification purposes for which they are designed.

#### 18. Significance

Test measurements are meaningful. The score on an Army test is not merely a number that is large or small, high or low. It is, first, a number

that signifies the amount of a specified attribute which a man possesses. According to the way in which the test is constructed (see ch. 3), the score will indicate an aptitude for a given training course or for a given assignment. And secondly, the score is a number that signifies how much of the required aptitude the soldier possesses in comparison with all other men in the Army.

### 19. Economy of Time and Ease of Administration

a. No matter how objective, reliable, and significant measurements may be, they are of little value to the Army unless they can be obtained quickly and easily. An Army is engaged in a constant race against time. Decisions cannot be delayed for weeks or months, but must often be made in a matter of hours. Most Army tests are of the paper-and-pencil variety that can be given to hundreds of men at one time. Also, most of the tests used by the Army can be scored quickly and accurately by means of the scoring machine.

b. The tests developed by the Army are different from most other tests in one important respect. Because the selection problem is not limited to any one place or type of installation, it is impossible

to procure enough men previously trained as personnel experts or examiners to carry out the whole program. As a consequence, much of the Army testing is of necessity done by men selected and trained for that work in the Army. In order to simplify the job, and to reduce to a minimum the demands made upon the individual judgment of the examiner, Army tests are so constructed that they do not require the services of highly skilled specialists in testing. The directions are uniform and explicit and conveyed in non-technical language. In fact, many of these tests are practically self-administering. Another advantage gained by making tests as nearly self-administering as possible is consistency in the way a test is given. In order to be fair to all men, the results of every test run must be comparable directly with the results of every other administration of the same test. To make test results comparable they must be given under the same conditions. The more individual judgment, skill, and invention play a part in administration, the greater the chance for variation. The Army has therefore left as little as possible to be decided by the examiner. Hence, the best examiner is the one who follows directions most closely and most intelligently.

## CHAPTER 3

### HOW THE ARMY CONSTRUCTS TESTS

#### Section I. PLANNING TEST

##### 20. General

Psychological tests are the measuring instruments of the classification system. They serve the same general purposes as physical tests, enabling the Army to weed out the completely unfit and to select for various assignments on the basis of characteristics which make men likely to succeed. Moreover, they predict how well any man will perform assignments in training or in the field. Every Army test makes it possible to observe the behavior of each man accurately and in exactly the same way as others are observed. Every test is capable of providing data which is accurate enough and reliable enough to result in a higher percentage of correct predictions than would be possible without it. Every test furnishes a method of recording the results in exactly the same way each time it is used. The results of every test may be interpreted in the same way by everyone who makes use of them. But no instrument is better than the people who use it. Therefore, all personnel in the classification system itself and all officers exercising a command function in regard to classification, assignment, or redeployment, require a thorough understanding of the nature of Army tests and the proper interpretation of scores. Such understanding can best be conveyed by a brief description of the principles and practice of test making.

##### 21. Tests Are Designed to Meet Specific Army Needs

The first step in test making is to study the classification problem to be solved. The particular need of the Army must be clearly defined both in terms of the assignment as a job of work or a course of study, and in terms of the number of men needed and the apparent supply. Only when the problem is urgent and important is the difficult and time-consuming process of test making warranted. If a rating scale, a questionnaire, or a survey of past experience will do just as well, or if the number of men and jobs involved is small, the construction of a special test is not warranted.

The test-maker, as will appear later, must continually bear in mind the practical purpose to which the test will ultimately be put. Moreover, study of the problem also brings to light the characteristic which the test should measure. It is necessary to find a characteristic which is possessed in high degree by most (if not all) of the men who have demonstrated their ability in a particular course or assignment. By measuring the amount of this trait possessed by untried men, it is possible to predict their performance quite accurately. In some cases, it is a very simple matter to find a trait highly correlated with success—successful carpenters possess carpentry skill, and it was no great feat to discover that this was the trait to measure in order to select the Army's carpenters. In other cases, especially highly complex and modern assignments, the particular traits must be discovered through experiments made after the general purpose of the test is decided upon. For example, it was decided to measure "cryptography aptitude" in order to predict which men would be most likely to pass a course in cryptography satisfactorily. But it took considerable research to find which particular traits make up cryptography aptitude and choose those most highly correlated with successful performance in the course. (See par. 32c.) It is essential to select for measurement traits which are actually to be found in the men themselves and not in the circumstances of a training course or assignment. If, for example, men pass or fail a course because of the whim of the instructor or the season of the year, there is no use in devising a test to discover the most likely candidates.

##### 22. Suiting Test to Its Purpose

The use to which a test will be put determines the kind of test to be developed and the form it shall be given.

a. Achievement tests are used for the purpose of finding the men who, without further training, will be most likely to succeed in a particular assignment. Achievement tests are, therefore, most often used in selecting men for direct assignment to some tactical or service organization. Achieve-

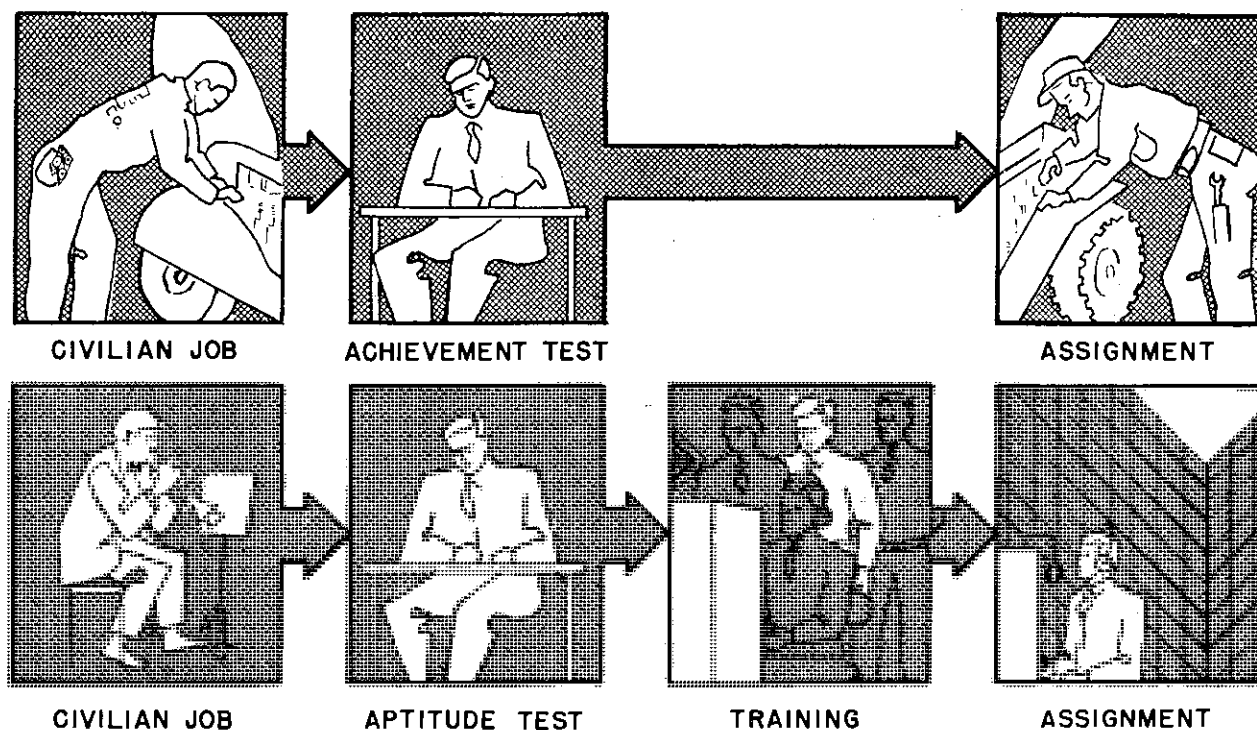


Figure 5.

ment tests predict by discovering and measuring what the examinee already knows, or how skillful he has become, or both.

b. Aptitude tests predict which men will be most likely to complete a training course in a minimum time and in a satisfactory manner. They are, therefore, most useful in selecting from among a mixed lot of available soldiers the most promising trainees for a particular course. They predict success by measuring the degree to which examinees possess traits found in men who have been successful in the particular training regime for which selection is being made.

c. Aptitude may be measured by an achievement test, provided that the possession of certain knowledge and/or skill indicates aptitude for acquiring related knowledge and skill. For example, the Army carpentry test can be used to choose men to be trained as general utility repairmen, because carpentry skill indicates *aptitude* for this specialty.

d. Army tests are as specific as may be necessary to accomplish a particular purpose. For example, it is necessary for general classification purposes to predict probable success in a wide variety of assignments in which general learning ability (sometimes defined as "intelligence") is the most

important factor. Consequently, the Army General Classification Test measures a wide variety of knowledge and ability such as most people acquire, to greater or lesser degree, in civilian life: ability to read, knowledge of word-use, arithmetic, and so on. The Mechanical Aptitude Test is more limited in range, but still far from being completely specific inasmuch as it measures aptitude associated with success in any one of a variety of mechanical assignments. The Army Radio Code Aptitude Test is highly specific, since it predicts the men most likely to succeed in a very specialized and homogeneous group of assignments, and measures only the traits which are correlated with performance in these assignments. The Army thus has a wide range of tests, from the highly general to the highly specific, each of them adapted to a somewhat different classification problem. In constructing a new test, the Army psychologists choose general or specific material according to the use for which the test is intended.

### 23. Form of Test

Two principal factors determine the form in which a test is cast: the purpose to be served and practicability. A test is a field instrument and must be designed to serve efficiently under conditions likely

to be found in the field. It must be adaptable to Army necessities which limit the time that may be allotted to testing.

a. **VERBAL FORM OF TEST.** A verbal test is one in which the examinee is required to talk, write, or mark correct responses stated in language. The Officer Candidate Test (OCT-1 and -2) is a typical example. The type of verbal test most commonly used is a paper-and-pencil test administered to groups. A verbal test of this kind can be administered and scored in a short time and with great efficiency and does not require the presence of a highly trained specialist. Test situations can be made uniform, and highly objective standards of scoring may be employed. A wide range of ability and knowledge may be sampled in a relatively short time. When a large number of men is to be tested, or highly trained personnel is not available, a verbal test is likely to be the most practicable, unless it is clearly unsuited to the purpose at hand. It should be borne in mind that a verbal test may measure either achieved skill or aptitude. A man who can answer certain questions concerning a job, or comprehend selected written passages about it, can usually do the job or learn to do it.

b. **PERFORMANCE FORM OF TEST.** A performance test is one in which the examinee is required to manipulate objects, deal with visual materials such as pictures and patterns, or make practical application of knowledge. It is the most efficient form to use when it is necessary to observe *how* the examinee does the job as well as his ability to do it. Performance tests involving an actual work sample, as for example the making of a mortise-and-tenon joint as a test of carpentry skill, bear an evident relation to the assignment and are both practical and interesting. The practicability of performance tests is limited by the fact that they are time-consuming and usually require highly trained personnel for their administration. They are, however, indispensable in testing illiterates and men whose knowledge of English is limited. Verbal directions may be very simple and require no response in language. The Group Target Test (GT-1) is an example of the performance test in which language is minimized in order to discover the aptitudes of men deficient in the use of English. Nonlanguage tests require no speaking, reading, or understanding of language on the part of the examinee in connection with either directions or response. They are used

almost exclusively to test men who do not understand English.

c. **GROUP TESTS.** Verbal (pencil-and-paper) tests are widely employed by the Army because they are best adapted to group testing. The large number of men to be classified and the pressure of time make it necessary to test in groups whenever possible. Performance tests, by their very nature, are better adapted to the testing of individuals. It is sometimes possible, however, to devise a performance test to be given to groups; an example is the Group Target Test, employed in induction stations. (See ch. 7.)

d. **INDIVIDUAL TESTS.** Individual tests may be of either the verbal or performance form, depending upon the particular purpose to be achieved, or a test may involve both performance and verbal responses, as in the case of the Army Individual Test (AIT-1). Individual tests are constructed to accomplish purposes for which group tests are not suited. These purposes are:

(1) Screening and classifying men whose language deficiencies or other personal characteristics render them unable to demonstrate their abilities adequately on a group test. The Individual Target Test (IT-1) is an example.

(2) Testing aptitudes or proficiencies which can best be revealed through actual work samples or manipulation of performance materials. The Distributor and Valve Test (TC-15a) is an example of this type of test

## 24. Items

Having decided upon the trait or traits to be measured and having determined the form which the test is to take, the psychologist's next step is to determine which, among the several possible kinds of questions or problems, will most effectively reveal the desired information about each man to be examined. Questions and problems are called *test items*. In performance tests, the items are usually problems involving cards, blocks, tools, materials, etc. In paper-and-pencil tests, the items are questions of two types:

a. **FREE-ANSWER ITEMS.** These may be sentences containing a blank space in which the examinee is instructed to write a word which makes the statement complete and correct, as in the following example:

The capital of ————— is Boston.  
Free-answer items may also be questions which the examinee is required to answer by writing a

word or phrase or sentence, as in the following:

In the Diesel engine, how is the gas mixture in the cylinder ignited?

There are situations in which it is desirable to discover not merely whether the soldier can recognize the solution to a problem but how he arrives at the solution or how he expresses himself in giving the answer. Where this is the aim of the test, the free-answer item is employed. It should be recognized that in giving credit for answers to this type of item, much will depend upon the personal judgment of the scorer as to what constitutes a good, an acceptable, or an entirely unsatisfactory solution or expression. To avoid these difficulties, "restricted answer" items are employed when practicable. The commonest form of this type is the multiple-choice item.

b. **MULTIPLE-CHOICE ITEMS.** Multiple-choice items present the examinee with several (usually four or five) answers to a question. His problem is to choose and write down the correct answer.

Examples:

Boston is the capital of—

- A) Maine
- B) Montana
- C) Massachusetts
- D) Minnesota

In the Diesel engine, the gas mixture in the cylinder is ignited by the—

- A) Spark
- B) Heat generated by compression
- C) Ignition system
- D) Firing order of the cylinders

Multiple-choice items are preferred for most testing purposes for the following reasons: Scoring is more objective because the right answer is already set down and is neither arguable nor subject to the varying judgments of testing personnel. The examinee has only to recognize the answer, and is not burdened by having to search for it in his mind and then phrase it in his own way. Because the examinee does not have to write, merely being required to check the correct answer, he can cover many multiple-choice items in a given time. As will appear later, it is an advantage to cover a considerable number of items. Multiple-choice items can be scored by machine, which increases accuracy and saves time. (See ch. 4.)

## 25. Length

To include all the items pertinent to a given trait would make the test absurdly and inefficiently

long. The principle which governs the number of items selected—and therefore, the length of the test—is that there must be enough to show the degree to which *each* examinee possesses the trait, and show this in a measurable fashion so that men can be compared with one another in terms of the trait. Classification testing employs the same sampling principles followed in other fields of measurement. In grading a carload of wheat, for example, it is not practicable to examine the whole lot in order to compute the percentage of high-quality grain, of chaff, and of foreign materials. The examiner instead gathers samples, assays these, and assumes that the characteristics of the whole carload are the same as those for the samples which he tested. But he would be extremely naive if he took all his samples from the top of the car. The unscrupulous vendor could easily have filled the car with an inferior grade of wheat and placed a thin layer of first class stock on top. The sample taken from this top layer would not be representative of the whole, and the measurement based on this sample would be an exceedingly inaccurate determination of the quality of the entire lot. Aware of all the pitfalls of careless sampling, and wishing his sample to be representative of the whole carload, the examiner collects a number of smaller samples—from the top and bottom of the car, from different depths, from each end, and from the middle. The more samples he collects, the more accurate his grading, since with only a few, chance discrepancies loom large in the total. The car may contain a concentration of inferior wheat, constituting a very small fraction of the total amount. If the examiner takes only five samples, and happens to take one of them from this small concentration, the inferior grade will comprise one-fifth of his total sample, and measurements based on it will not be characteristic of the whole carload. So with testing, the larger the number of items, the greater the accuracy of the test—for two reasons: the larger number provides more complete coverage of the whole content and at the same time insures that any small pockets of ignorance on the part of the examinees will not be given disproportionate weight in the final result. There is a second consideration, a practical one, that enters into the determination of length of the test. The Army cannot afford to devote an excessive amount of training time to the administration of tests. Furthermore, when a test becomes overlong, the effects of fatigue and boredom are apt to diminish the accuracy of the



results. In practice, the test-maker includes as many items as are necessary to make the test sufficiently accurate for sound classification, but keeps it short enough to be practical under field conditions and within the physical capacities of the average man.

### A GOOD TEST SAMPLES ALL PARTS OF THE SUBJECT MATTER FIELD

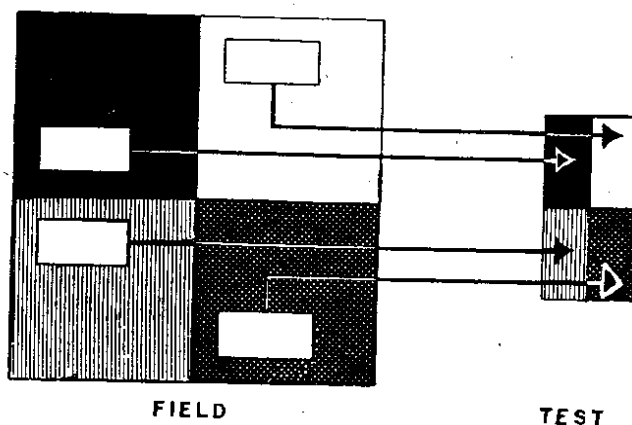


Figure 6.

## 26. Time Limits

Such limits are extremely important because they help to determine the qualities measured by a set of items. A test made up of items which are equal in difficulty, measures *speed* if the time limit is so short that no one can finish all the items. A test in which the items get harder and harder measures *power* if examinees are given all the time they need to complete as many items as they possibly can. Most Army tests measure both *power* and *speed*, since war neither waits upon the slow-but-sure nor makes allowances for the lightning blunderer. The Army must find out both how well and how fast a man can be expected to perform in a given assignment. The Officer Candidate Test is a good example. It is made up of questions and problems which become harder as the examinee forges through them. Only the man who is both able and quick can complete most of the 70 items in the 45 minutes allowed to the test.

## Section II. CONSTRUCTING TEST

### 27. General

Having "blueprinted" the test by making basic decisions as to the trait to be measured, the form

and character of the test, its probable time limits and length, the next step is to build a model which can be given a thorough trial. The actual building consists of writing items and preparing directions for administering and scoring the test.

## 28. Construction of Test Items

The first step is to assemble a large collection of questions or problems which are related to the trait to be measured. When necessary, an expert consultant with a highly specialized knowledge of the subject matter is called upon to aid in getting this material together. The next step is to frame the questions or problems into items; this highly technical and difficult task is performed by psychologists who have made item-writing a specialty. Each item is checked to make sure that it is in the proper form, clearly phrased, pertinent, and relevant, and calculated to elicit a response which indicates presence or absence of the trait which is being measured. The entire collection of items is analyzed to make sure that each item contributes to adequate coverage of the field.

## 29. Preparing Directions

a. The directions which accompany each set of test items constitute a statement of the conditions under which the test was "calibrated" or standardized. Great pains are taken to make these directions complete and clear. Only by following them can the standard conditions—the conditions under which the test was standardized—be repeated. Unless these same conditions prevail, a test gives results as unreliable and misleading as a thermometer reading taken when the patient has a mouth full of ice.

b. Two sets of directions are prepared for each test. Instructions and suggestions to the examiner are included in the manual that accompanies the test whenever it is administered. They indicate the general conditions under which the test should be given, list the materials required to give it, and the time limits for the parts and for the whole test. They also suggest introductory remarks that should precede, and set the stage for, the administration proper, as well as answers to questions that commonly arise during the testing session. The second set of directions are the specific instructions to the examinees. These are printed as part of the test booklet itself to insure that instructions will be the same for every administration of the test. Their purpose is to make certain that each individual

examinee understands just what he is expected to do and how he is expected to do it. They touch upon such details as the advisability of guessing when not absolutely sure of the answer, the amount of time that will be allowed for the test, the relative importance of working for speed as against working for accuracy. And they give precise and detailed explanations, along with demonstration and practice items, of the correct manner of indicating answers. Since all of these directions constitute such a vital part of the test, they are prepared by experienced test psychologists and subjected to independent check for completeness and clarity.

### 30. Preparation of Scoring Directions

a. GENERAL. The final step in making a model of a new test is to work out the proper technique for scoring. The items are first put in the order which experts believe will yield the best results. The right answers are given a final check to make sure that they are clear and matched with the questions to which they belong. The position of the right answers is adjusted so that they fall in truly random positions. That is, the right answers are so located that the examinee will not be able to "outguess" the test by using or discovering a particular pattern of right answers. The truly random arrangement makes certain that all the correct guesses made by the examinee will not be greater than those obtained by pure chance. "Randomization" is accomplished by making a "scoring key" or adopting one of the standard scoring keys furnished with International Test Scoring Machines. A scoring key indicates the position of the right answer on the answer sheet. The correct answer is placed in the identical position in the test booklet. The incorrect alternatives are so arranged as to reduce to a minimum or eliminate entirely any clues that the examinee might derive from the sequence of the alternatives.

b. THE SCORING FORMULA. (1) It has been found advisable to take further precautions in scoring tests where guessing may give a man a higher rating than his abilities warrant. The following example makes clear both the precautionary technique itself and the reasons for applying it. If an examinee selects one of four alternatives of a multiple-choice item by pure guess, his chances of

selecting the correct alternative are one in four. In a large number of such guesses, he will be wrong three times for every time he guesses right. On a 100 item test, for example, he will usually obtain 25 right choices by answering in this fashion. Since he got one right answer for every three wrong ones, his "true" score can be obtained by the simple process of subtracting from the number he got right one third of the number wrong, according to the formula:

$$R - \frac{1}{3} W = \text{"true" score.}$$

Applying the formula to this case (25 minus  $\frac{1}{3}$  of 75) will give the examinee's correct score. One further example will illustrate the technique. Let it be assumed that two examinees each know the answers to 50 items of a test, but that whereas one of them stops at this point, the other goes on to make pure guesses on the next 20 items and, by chance, gets five of them right and fifteen wrong. The obtained score of the first examinee will be 50 and that of the second, 55. Application of the scoring formula to both cases, however, will give the first man (50 minus 0) 50 and the second man (55 minus  $\frac{1}{3}$  of 15) also 50. It is important to note that the fraction in the formula depends upon the number of alternatives to each question.\* For a test composed of items having five answer choices, the formula would be right minus one-fourth wrong.

(2) The use of the scoring formula is based on the logic of chance. In practice, however, guesses are seldom completely blind. An examinee may get an item right by knowing which alternative is correct or by knowing that the other three are wrong. Likewise, if he knows that two are wrong, he will have to guess only between the remaining two and will, therefore, stand a better chance of picking the correct one. Any error that results from the application of the correction formula will always be in favor of the examinee who utilizes such judicious "guessing." However, there will be other, more cautious examinees who may know just as much but who will never put down an uncertain choice if they are to be "penalized for wrong answers." It has been found that the correction formula is not sufficiently helpful in estimating the probable success of examinees to justify the additional work involved, except in certain tests of general ability and tests where pure guessing is common. It is, therefore, used only with such

\*The fraction is always  $\frac{1}{n-1}$  where n is the number of alternatives to each item.

tests as the Army General Classification Test, the Army Radio Code Aptitude Test, and a few others.

### Section III. TRYING OUT THE TEST

#### 31. General

After the planning and construction of the experimental model of the test, the next step is the "shake-down." The aims of the trial and analysis of the new test are, in general, the same as those for any other shakedown. A new gun is tested to make sure that it will shoot accurately and consistently and according to its specifications. A new test is given a trial to make certain that it will measure specified characteristics which determine the classification of soldiers.

#### 32. Field Trials and Analysis

In the experimental model of a test, there are many more items than will be used in the finished product. On the basis of the findings obtained in field studies, certain of these items are rejected and the others rearranged to make up the test in its final form.

a. **TEST POPULATIONS.** As a first step, the experimental form is administered to men who are representative of the soldiers who will be classified by means of the same test in finished form. Tests which will be given at reception centers are taken to a typical reception center for trial. A test to be used for screening candidates for officer training is given to groups of men who meet all other requirements for selection, as is the case with the entering classes in officer candidate schools. It is of the utmost important that the group on which a test is tried out be a truly representative sample of the group to be classified by means of it. If items destined to form part of a test for *all* reception center men are administered in the "shake-down" stage to a single group arriving from sub-marginal communities in the hinterlands, no clear or dependable indication can be gained of the effectiveness of these items in measuring the traits of men from other, more fortunate sections. If a test is to be used with all men at training centers to select for specialist training, it must not be given experimentally to trainees already enrolled in the course. A group is said to be representative of its parent population if all its various characteristics are the same as those of the larger group from which the men to be tested are drawn or, in more technical terms, if each member of the parent population has

an equal chance of being included in the sample group. These conditions are obviously not present in the cases cited above. In the training center example, all trainees would not have the same chance of being included in the specialist group, since this class has already been selected in some fashion, and would, therefore, rate higher on those traits for which they were selected than would the whole training center group.

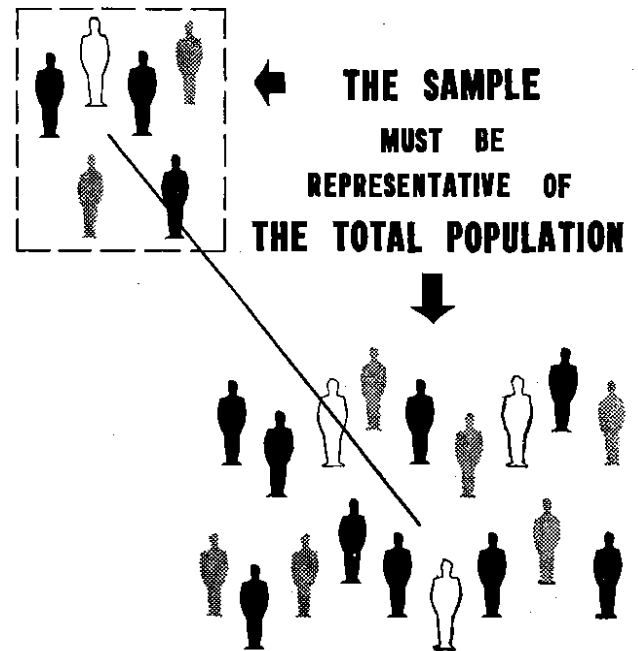


Figure 7.

Extreme care is exercised to make sure that the sample group to which the test is administered has all the characteristics of the parent population. Whenever practicable, the desired representativeness is achieved by testing a truly random sample. Representativeness is achieved in other cases by selection controlled with respect to general intellectual capacity, age, color, education, and any other characteristic that might be related to performance on the test items.

b. **DIFFICULTY AND DISCRIMINATIVE INDEX OF ITEMS.** The experimental model of the test, then, is administered to a group representative of that for which it is designed, and the results analyzed to determine the difficulty and "discriminative index" of each item. Difficulty is not determined by the subjective estimate of the test-maker, nor the consultant, nor any opinion that the item "should be easy for anyone claiming to be familiar

# A DISCRIMINATING ITEM DISTINGUISHES BETWEEN THOSE MEN WITH LOW SCORES AND THOSE WITH HIGH SCORES ON THE TOTAL TEST

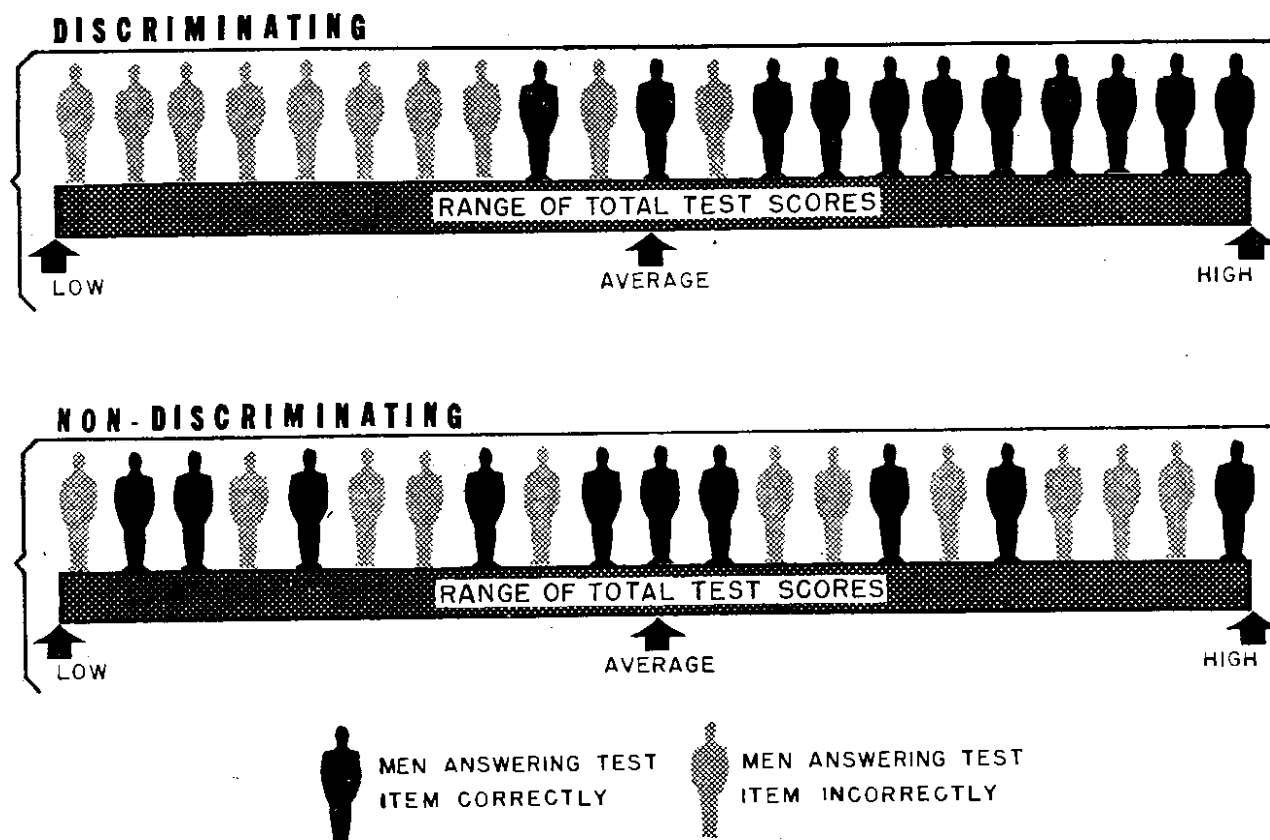


Figure 8.

with the content of the test." The difficulty of a test item is proved by facts; that is, by the proportion of men in a representative group who actually do answer the item correctly. The computation of item difficulty, therefore, involves the simple but tedious task of counting, for each item, the number of examinees who received credit for that item during the experimental run of the test. Difficulty is expressed in percentage form; a difficulty of 70, for instance, means that 70 percent of the group answered the question correctly. This would be a relatively easy item. An item having a difficulty of 28, that is, an item answered correctly by only 28 percent of the group, is relatively hard. The term *discriminative index* is used by the Army to designate the value a particular item has in ranking men according to the varying amounts of the tested trait which they possess. If, for example, the individuals who answer a given item

correctly tend to answer most other items correctly, while those who fail on the item receive low total scores, then the item will be contributing to the differentiating function of the test. It will have a high discriminative index. An item which, regardless of its difficulty is answered as often by those who fail most of the other items as by those who get high total scores, does not add to the ability of the test to indicate the differences between men. If an item is answered correctly by all the examinees, it merely adds the same amount to the score of each individual without in the least affecting their relative ranks. The same is true of an item which no one can answer, and true also for an item which is answered by the same proportion of low-ranking and high-ranking men.

c. **CRITERIA AND VALIDITY.** No *a priori* or subjective opinion will tell with enough certainty for Army purposes whether an item actually has

anything to do with the trait to be measured. Evidence in large quantities is required. To get that evidence, a criterion against which the item itself can be measured must be established. Here the psychologist's problem is more complicated than that of the maker of familiar measuring instruments. The length of a yard has been agreed upon for a long time; therefore, a precise yard kept under constant temperature at the Bureau of Standards is a criterion for all yardsticks. The psychologist must measure his test against actual performance. An item is known to be *valid* when there is acceptable proof that it has value in predicting the particular performance taken as a criterion. Success in some assignments requires that men have a number of well-defined characteristics. For example, men who are proficient in cryptography are able to encode and decode at satisfactory speed, know various codes, and make sound and rapid deductive judgments. All of these are criteria of cryptography aptitude. When it can be proved that an item helps to predict the likelihood that a man will demonstrate any of these phases of cryptography aptitude, the item is in accord with the criterion and therefore valid for a cryptography test. The Army tests items against criteria by making a second tryout of the test.

**d. RELATION OF CRITERIA TO ACTUAL PERFORMANCE.** The experimental items are administered to men who are about to enter a course or enter upon an assignment. Subsequent performance of these men is carefully observed. All significant phases of their performance are recorded so that each test item can be compared with the most complete measure of the criterion and its validity thus determined. For example, the speed in encoding and decoding achieved with the various cryptographic devices by men in the cryptography course could be used as a criterion. Only items answered correctly on the test by a high percentage of men who later achieved satisfactory speed, but answered incorrectly by a high percentage of those who failed to achieve such speed, would be considered valid in and of themselves. Items selected by this criterion would predict only this phase of cryptographic aptitude. To predict all phases—speed, knowledge of various codes, methods and deductive reasoning—a composite criterion is used.

### 33. Selection of Test Items

The field studies and analysis of the experimental form of the test will have furnished data on the three main characteristics of each test item—its difficulty, its discriminative index, and its validity. The next step is to select the items which will make up the final form of the test in such manner that the finished product will be a carefully calibrated, reliable, and valid instrument. This selection is always based on the three characteristics of the item mentioned above, but it should be noted that these characteristics may be wholly unrelated to each other. An item of high validity may be difficult or easy, and it may have high or low discriminative value. It is accordingly impossible to select on the basis of any one characteristic at a time. All three must be taken into account.

**a. LENGTH AND DIFFICULTY.** The number of items at each level of difficulty is determined by the purpose of the test. If this purpose is to make the most efficient division of the population into a high and low group, with reference to the trait in question, the difficulties of the items selected cluster around the division point. More specifically, if it is desired to qualify the top 30 percent of a population for specialist training or assignment, then the difficulties of the items selected should cluster around 30 percent (items answered correctly by 30 percent of the population). If however, as is usually the case, it is desired to grade the whole population from highest to lowest with reference to a trait, rather than merely to divide into two groups, the difficulties of the selected items should be spread over most of this range. In most tests the item difficulties will be fairly evenly distributed over the range from 30 percent to 70 percent.

**b. DIFFICULTY AND DISCRIMINATION.** In paragraph 32b, it was suggested that item difficulty and discriminative index are interdependent values to the extent that a very easy or a very hard item cannot have a very high discriminative index. While this is true for the whole range of ability covered by the test, such an item may discriminate well over a narrow range. A very difficult item, for example, might be failed by all of the lower four-fifths of the examinees, but if it is answered by as many as one quarter of the highest scoring fifth, it will be a very valuable item. For it is necessary to select items which discriminate at all levels of difficulty. Figure 9 illustrates good and bad discrimination at various levels of difficulty.

# SHOWING GRAPHS OF GOOD (HIGH DISCRIMINATION) AND POOR (LOW DISCRIMINATION) ITEMS AT THREE LEVELS OF DIFFICULTY

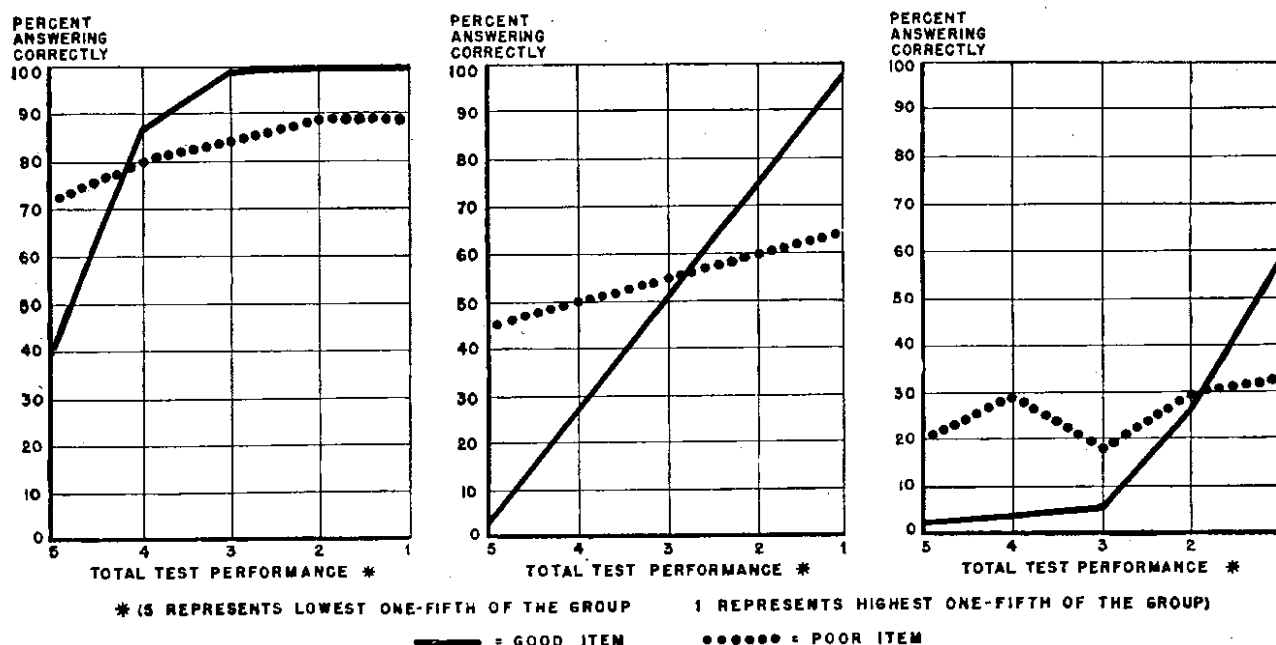


Figure 9.

In the first chart, for example, are graphs representing a good and bad easy item. Both have about the same difficulty. But the graphs make it clear that the good item differentiates rather sharply between the poorest group and those only slightly better, while the bad item is answered by good and poor with approximately equal frequency.

c. **FINAL SELECTION.** Items are sorted according to their difficulty and their discriminative value. Final decision to accept or reject any item is made on the basis of its validity. Where several items are approximately the same with respect to the other two constants (difficulty and discriminative index) the ones with the highest validity are selected. Of course, items which have no validity regardless of their other characteristics, are invariably rejected. Items which have the same discriminative value and the same difficulty duplicate one another and are therefore so much dead weight. In such cases a very slight difference in validity is basis for rejecting one or more items.

## 34. The Final Form of Test

Having selected from the collection of items those which will serve best the purpose of the test, the total reliability and validity of the finished product

is determined by rescoring the answer sheets for these selected items. Except for certain minor considerations, this rescoring will give the score that each individual would have received had he been given the final test rather than the experimental model. Consequently, by rescoring the papers for the first tryout of the items with a representative group (see par. 32a), and by performing the appropriate statistical operations, the reliability of the final test can be computed, and, if not satisfactory, adjusted by the addition or reselection of items. By rescoring the answer sheets for the second tryout, with a population for which a criterion is available (see par. 32d), the validity of the total test in final form can be computed.

## Section IV. ESTABLISHING THE SCALE OF MEASUREMENT

### 35. Purpose of Standardization

Upon completion of the item selection, the test is a finished product in the sense that it is an accurate instrument capable of producing dependable measurements. But these measurements will still be in terms of "raw" scores, that is, the number

of questions answered correctly, or the number right minus a fraction of the number wrong. By itself, a raw score is seldom of much value to the classification officer who has to use it, regardless of how accurate and dependable the test may be. A raw score does not tell the degree to which a man possesses a given skill or aptitude *in comparison with other men in the Army* and is, therefore, no clear indication that he will do better or worse than others on assignment. A raw score does not tell what proportion of Army men stand higher or lower in regard to the trait under consideration. For each test, it is necessary, therefore, to know the scores of all Army men and how they are distributed along the range from high to low. Classification officers have neither the time nor the facilities for collecting this necessary data. Further, a time-saving and efficient technique for interpreting raw scores in terms of this data is required to make sound classification practicable in an Army of millions. The data concerning performances of all Army men is obtained by testing a *standardization population*, as described in paragraph 36. The device for handy interpretation by the classification officer is the *Army standard score scale*, developed by the test-makers to show what raw scores mean in terms of Army requirements. (See ch. 5.)

### 36. Standardization Population

It is neither feasible nor efficient to give each new test to the whole population of Army men. Sufficiently dependable information is obtained by careful sampling methods. Each new test is given in its final form to a large group carefully selected to represent the whole Army population as accurately as possible. The size of the group varies considerably, depending upon the nature of the problem, the availability of groups, and the requirements of speed and economy. No practical advantage is gained by enlarging the sample at a high cost in time, energy, and personnel, since scientific control of selection and the application of statistical techniques produce results that meet the requirements of sound classification. The representative group to which the final form of the test is given is the standardization population. The administration of the test to this population and the statistical computations which follow are known as standardization.

### 37. Critical Scores

The score below which men may not be accepted for an assignment or training course (or in the case of induction test for the Army itself) is called a *critical score*. Critical scores for several different assignments may be set at appropriate points on the range of scores for a single test. For example, a score of 110 or above on the AGCT meets one of the requirements for entry into Officer Candidate School, while a score of 100 on the AGCT is sufficient to admit a man to candidacy for several of the Army Service Schools. The chances of success indicated by any obtained Army standard score can be readily computed. The critical score is set at a point dictated by Army necessities. Thus, if it is desired that 80 percent of the men selected shall complete a course successfully, or perform satisfactorily in a given assignment, the critical score could be set such that only men who stand a 4 to 1 chance of success will be selected. To select so high, however, may also mean that few will qualify. In establishing the critical score, it is therefore necessary to take into account the supply-demand ratio for the particular course or assignment in question. If the demand is small in relation to the supply, the critical score can be set high and there will be less probability of failure among those selected as in the foregoing example. Where the demand is relatively large, however, it will be necessary to lower the critical score in order to qualify more men. When this is done, some of those who are selected will stand a smaller chance of success and consequently a higher percentage of failures must be expected. (See par. 69.)

### 38. Continuing Studies

Supply-demand ratios change from time to time, and even the nature of the course or assignment may be materially altered as a result of mechanical and technical innovations. The Army's psychologists must make sure that every test which is a valid instrument for predicting success in training also proves a sound predictor of later job performance. When the test is released for field use, it remains a concern of the test makers. They make follow-up studies from time to time, checking the subsequent job performance of men who scored high and those who scored low. Only through the study of accumulated evidence can the Army be sure that the test is doing what it is supposed to do, or that an improved edition is necessary.

## Section V. SUMMARY

### 39. The Steps in Test Construction

These are the steps in test construction and the major purposes of each:

*a.* Study of classification problem to determine need for test, practical considerations influencing test structure and trait to be measured.

*b.* Design of test or tests to accomplish purpose clearly defined in step *a.*

(1) Selection and evaluation of criterion.

(2) Selection of kind of test.

(3) Selection of form of test items.

*c.* Construction of items calculated to reveal potential competence of men.

*d.* First experimental tryouts to determine—

(1) Discriminative index of each item.

(2) Difficulty of each item.

*e.* Second experimental tryout to determine—

(1) Predictive value of each item in terms of criterion.

(2) Predictive value of criterion in terms of actual assignment.

*f.* Selection and sampling of items which have proved most effective.

*g.* Computation of validity and reliability of test as a whole.

*h.* Establishment of field form of test for actual use.

*i.* Standardization process—Construction of standard score and conversion tables for interpreting test.



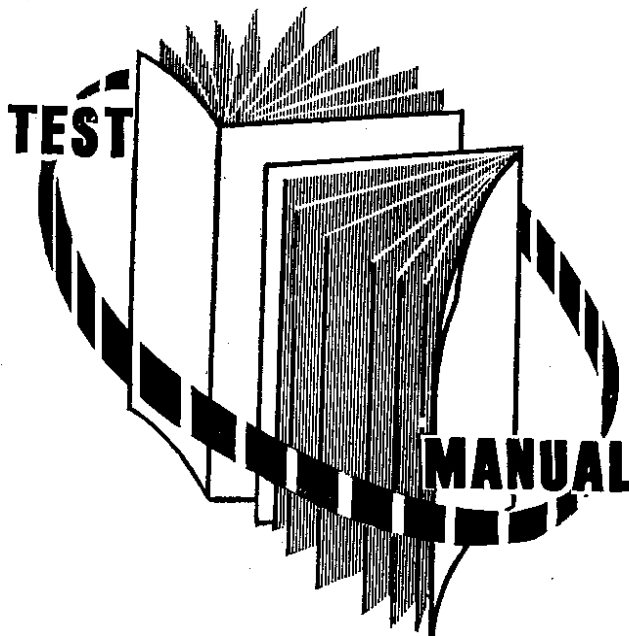
## CHAPTER 4

## THE ADMINISTRATION AND SCORING OF TESTS

## Section I. INTRODUCTION

## 40. General

The scientist develops tools, the technician puts them to use. The instruments devised by the scientist, constructed on the basis of extensive knowledge and calibrated with great pains, can be valuable means of increasing the accuracy of observations, insuring the objectivity of judgments, and revealing information. But the extent to which scientific tools are useful depends upon the discipline and the scrupulous care which technicians practice in using them. Like the physician's stethoscope, which reveals the condition of the human heart only to the trained observer, but is merely a gadget in the hands of a layman—so psychological tests are meaningless abracadabra unless employed by men who understand them and are willing to take the necessary pains. The aim of the present chapter is to clarify the work of the technician in giving and scoring classification tests so that they will produce results of the greatest possible value to the Army.



**THE MANUAL IS PART OF THE TEST**

Figure 10.

## 41. Importance of Testing Instructions

Specific directions for administering and scoring are set forth in the manuals which accompany each Army test and are as much an integral part of every test as the questions themselves. Included in specific instructions are the exact wording of the directions, the time limits of the test, the scoring criteria, and descriptions of the proper technique for recording and interpreting results. Deviations from any of these can affect the accuracy of measurement as much as counting wrong answers right. Directions must be adhered to strictly.

## 42. Testing Situation Must Be Standard

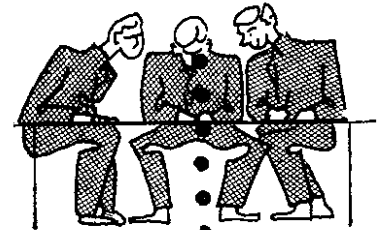
Since it is the function of every test to compare each individual with others in the Army population, it follows that the conditions under which tests are administered and scored must be the same for every soldier, regardless of when or where the test was given. The scores of men tested in noisy surroundings by slipshod methods are not comparable to those of men examined under favorable circumstances. Nor are they, in all probability, accurate indications of the real abilities of those men. The use of such scores can only result in improper classification and misassignment, with attendant loss to the Army of potential skills.

a. Testing conditions and procedures should be so standardized that if it were possible to find two individuals exactly alike, both would achieve the same scores, though tested at different times and in different places. Only scores obtained under standardized conditions can be relied upon to show what may be expected of men.

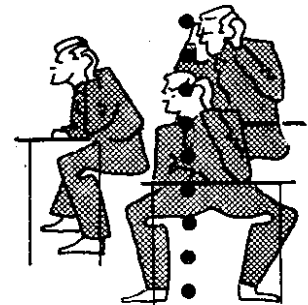
b. No valid comparisons can be made between the scores of men performing at different levels of motivation. The best method of approaching uniformity is to make sure that all men perform to the best of their ability. Standard conditions should therefore be optimal conditions.

c. Tests should be administered and scored in a manner identical with that employed in their standardization. Standardization (see sec. IV, ch. 3) involves the administration of each test to a population with known characteristics in order to obtain norms by means of which each subsequent

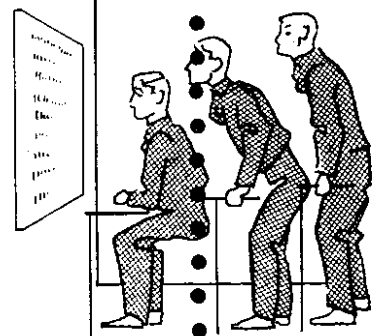
# GOOD TEST ADMINISTRATION REQUIRES CAREFULLY CONTROLLED CONDITIONS..



too curious

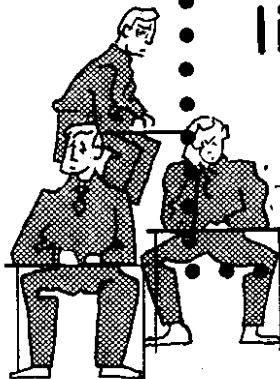


can't hear

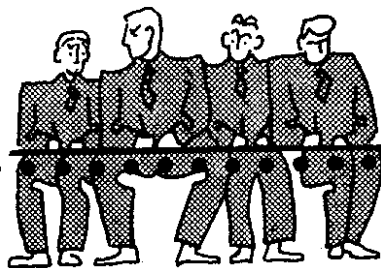


can't see

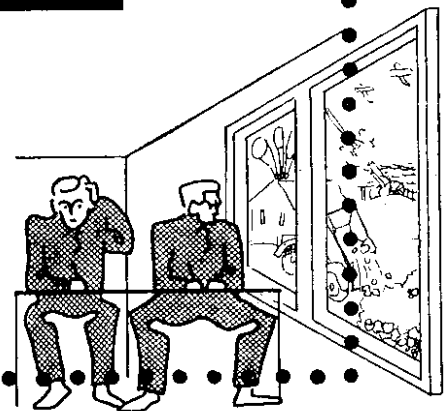
.. Avoid situations like these ..



too sick too tired



too close



distracting noises

Figure 11.

score may be evaluated and interpreted. In other words, test performances in the field are evaluated by comparing them with the performance of the men in the standardization population. If this comparison is to be a valid one, the administration and scoring should be identical in the two instances. Army tests are always standardized in the field under conditions which can be duplicated in field installations. The principles set forth in this chapter make it possible to duplicate the standardization conditions.

## Section II. PRINCIPLES AND PROCEDURES FOR ADMINISTERING GROUP TESTS

### 43. General

It has already been stated that the procedures for administering tests should be such as to call forth the best performance of which the individual is capable under standard conditions. Each individual will tend to do his best if his environment is reasonably free from distracting influences, if he understands what he is to do, and if he considers it worthwhile to do his best. The first of these conditions depends upon the physical aspects of the testing situation; the others upon the techniques employed by the examiner in controlling the testing situation.

### 44. Physical Surroundings

All behaviour, including test performances, takes place in an environment and is influenced by that environment. Since it is patently impossible to administer tests in a vacuum, the next best thing is to take steps to insure that the environment provided is standard for all administrations of the test, and that it does not impede or hamper the performance of the examinee. While it is recognized that ideal testing conditions cannot always be achieved with the limited facilities available in field installations, attention to the following factors should provide conditions that are adequate in most cases.

a. So far as possible, the testing room should be quiet. Noise is one of the principal sources of distraction from concentration and mental effort. Yet, absolute silence is neither necessary nor desired. The individual who can perform satisfactorily only in a soundproof room is going to find few places in the Army suited to his peculiarities. Noise which continues steadily at a moderate and fairly even level of intensity can be considered as

normal for testing conditions. Such noise would include the steady hum of indistinguishable voices from another part of the building, the drone of machines, or the continuous but muted clatter of typewriters. But a sudden shout outside a window, a bell, the clatter of unloading a truck, the blare of a radio, or the sound of persons passing through the room—these are stimuli compelling the attention of examinees and interfering with their test performance. The noise need not be loud to be distracting. The jingle of coins in a proctor's pocket or the cracking of nervous knuckles can be extremely upsetting.

b. Consideration should be given to the acoustics of the testing room. The examiner's voice must be clearly audible to all men being tested. The public address systems now found in most Army testing rooms have solved this problem, but not without adding others. Care should be exercised in placing loudspeakers and in locating microphones. The public address system is still something of a novelty, and people feel an urge to see where the voice is coming from. An invisible ghost-voice can cause considerable craning of necks and unnecessary distraction; so, if a test must be given from an unseen location, a preliminary announcement to this effect will dispel distracting curiosity. The level of amplification should also be controlled. Loud directions booming forth above one's head can be very disconcerting.

c. The testing room should be well lighted and ventilated. There must be sufficient illumination on the working surface to prevent eye strain. If a light meter can be obtained, the illumination in various parts of the room should be checked. Remember, illumination of the working space is the important thing. A light meter laid on this space should register approximately 6-10 foot-candles. Special care should be exercised to avoid glare spots and shadows; there is perhaps nothing as annoying as having part of the test paper intensely illuminated with the rest in the shadow cast by a pillar, a partition, or the examinee himself. Conditions of temperature, humidity, and ventilation are sometimes difficult to control, yet every effort must be made to do so. No one can perform at his maximal efficiency in a room where the air is hot, sticky, or stale.

d. Among the foremost factors of importance are the spatial arrangements of the testing room. If conditions permit, the examiner should be provided with a raised platform or rostrum in a part

of the room where he can see, and be seen, by all men being tested. This is especially important where test directions call for the presentation of charts or other demonstration material. The desks or tables for the examinees should be arranged to leave aisles for the proctors to use in distributing and collecting test materials and in circulating about the room during the test. If possible, there should also be enough space between rows to allow passage. Otherwise, when it becomes necessary for a proctor to reach an individual in the middle of a row, there is much treading on toes and knocking elbows en route. The distracting influence of such stimuli needs no further comment. The working space itself should be flat and smooth and free from cracks. Pencils have an irritating way of punching through the answer sheet when there are knotholes and cracks in the board beneath. If available tables are rough, a tight covering of linoleum or pressboard (masonite) will correct this. The space allotted to each individual must be wide enough to accommodate both a test booklet and a separate answer sheet. Chairs with writing arms should not be used for testing since the writing surface provided is far too narrow. Many installations now are equipped with large tables with vertical partitions separating the surface into booths approximately 30 inches wide and 18 inches deep. The partitions insure each person sufficient room and prevent the overcrowding of timid souls by neighbors with aggressive elbows. They also discourage community collaboration. If tables like these are not available and cannot be constructed, mess tables make an adequate substitute. However, if these mess tables are still being used for eating, the hours just before and just after meals should be avoided in the testing schedule. The noises and odors issuing from the kitchen or the clatter of dishes from another part of the mess hall fall into the category of distracting stimuli.

*e.* The temptation to give or to receive aid always seems to be present wherever people are examined en masse. Aside from the fact that cheating is reprehensible from the viewpoint of military discipline, its effect on the validity of the test score requires that it be prevented. For classification purposes, the Army is interested in how many correct answers the individual can obtain by himself, not how many he can copy from his neighbor. The use of partitioned booths (described above) or of alternate seating will help to prevent collaboration. In addition, all black-

boards and charts in the room should be checked to insure that no material is left visible to help the examinee, and all test booklets which are to be re-used should be examined after each session and any answers written therein erased. Despite all precautions, the proctors will still have to prevent cheating during the examination. For this reason, proctors should circulate (as quietly as possible) rather than remain at a fixed post. The mere nearness of the proctor on his rounds is often a sufficient curb on cheating.

*f.* Not all distracting influences are in the external surroundings. The condition of the individual, his physical and mental state, also affect his test performance. The man, for example, who has just had disturbing news from home, or is in physical distress, is in no condition to do his best on an examination. In individual cases, these factors cannot always be foreseen, but for the group as a whole, much can be done by scheduling testing sessions at a time of day when fatigue or physical or emotional discomfort can be expected to be at a minimum. In normal circumstances, the morning hours will be the best time to schedule an examination and the end of a long day the poorest. Where possible, activities should be controlled so as not to interfere with testing schedules. In the reception center, for example, processing should be so regulated that testing does not follow hard exercise, long hours of waiting in line, or immunization "shots." In all cases the test officer, examiner, and proctors should be alert to the signs of genuine distress, and the affected persons should be excused until a more propitious occasion.

#### 45. Testing Session

The ideal testing session is a smooth-running, organized affair. Since its primary purpose is to obtain reactions to standard questions under standard conditions, the major portion of the time is allotted to taking the test itself. All other activities such as assembling and seating the men, distributing materials, giving preliminary directions, and collecting materials—are necessary adjuncts to the main event. Yet it is the management of these details which can make the session a smooth-running operation or chaotic confusion—an ordeal to both examiner and examinee. The secret of success is control. If the examiner is at all times master of the situation, he can keep things moving and organized. But if the examiner is uncertain and stumbling, if there are unnecessary

delays, the group will become restless and irritable. Control is best achieved by careful preparation and practice in all phases of the process—preliminary arrangements, test administration, and the collection and disposition of materials. The discussion that follows will cover general principles and specific suggestions for making the testing session an orderly, systematic affair.

a. **PREPARATION.** Preliminary planning for the testing session involves the careful selection of the testing team, the instruction of all members, and practice drill in all required testing procedures. The examiner is selected for the quality of his speaking voice and for his ability to handle groups of men. While no one demands that the examiner have the speaking voice of a trained actor, he should have one that can be understood easily. His accent should either be indigenous to the group being tested; or, if the group is from many parts of the country, his accent should be "standard American"—the accent common to the Midwest. It is also desirable that the examiner be capable of controlling the testing situation—in many cases, noncommissioned grade gives the examiner the prestige necessary.

(1) The examiner should make a careful study of the manual to make sure that he knows the purpose of the test, the materials needed to give it, the directions to be read, and the problems which are likely to arise. He should study those directions which are to be read aloud until he can read them in a normal manner, without stumbling over unfamiliar words, losing his place, dawdling, or racing through in an unintelligible patter. Familiarity with the contents of the test itself is also invaluable. It is excellent practice for both the examiner and all proctors to take each test in the normal fashion before attempting to administer it; this procedure should be standard whenever a new test is installed or new examining personnel are trained. In this way, the examiner gains an appreciation of the men's viewpoint on the test and learns how to anticipate, and thus be prepared for, the common questions which may arise.

(2) The examiner is responsible for instructing the proctors thoroughly in their specific duties. The common practice of snatching any man not at the moment occupied, and making him a proctor, then and there, should be frowned upon. It is far more efficient to designate regular testing teams responsible for the administering, proctoring, and scoring of tests. Each proctor should be assigned

a certain section of the room for which he will be responsible. Before the testing period, he should check the materials to be used to make sure that they are all there in good condition and order, and in sufficient quantities. He should know the order in which these materials are to be distributed and collected, so that, when the time comes, he can execute this phase of his assignment with efficiency and dispatch. With the administration of the test itself, his real job begins. While directions are being read and while the test is being taken, he should patrol his assigned area. Within this area, he is responsible for:

(a) Seeing that each examinee has all the necessary materials for taking the test, and furnishing these, especially pencils, where needed.

(b) Insuring that each examinee is following the directions correctly and understands what he is to do and how he is to do it. The proctor should be alert to detect incorrect methods of marking answers where separate answer sheets are employed.

(c) Seeing that each examinee is doing his own work, independent of his neighbors.

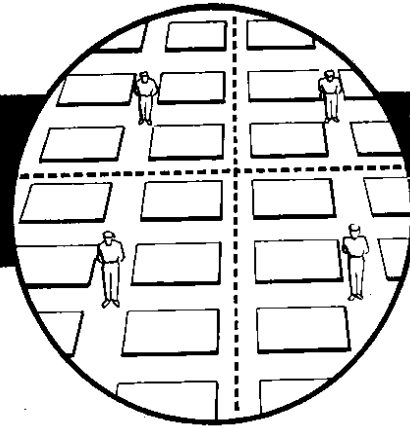
(d) Excusing from the examination any person who is or becomes too ill to continue without discomfort.

(e) Handling all inquiries of the men being tested. In no event, however, is he permitted to give information pertaining to the content or the meaning of the test questions, and he should inform the examinee that this is forbidden.

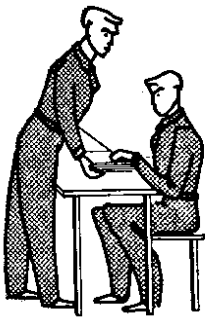
b. **ADMINISTERING TEST.** All Army tests should be administered strictly in accordance with the manuals of directions which are supplied with them. A test given without the aid of the manual is no more trustworthy and reliable than a rifle without sights. What follows here can be considered as instructions on how to use the manual in administering and timing a test.

(1) It has been previously said that the value of any test score will depend on the extent to which the examinee understands just what he is to do and the degree to which he considers it worthwhile to do his best. The examiner's primary responsibility, in fact his main function, is to elicit this willingness to work and to provide the proper instruction. The first is a matter of the appropriate stage setting and of favorable attitudes. The second is handled by oral directions contained in the manual.

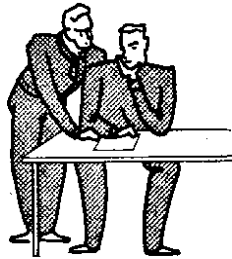
## EACH PROCTOR IS RESPONSIBLE FOR ONE SECTION OF THE TEST ROOM



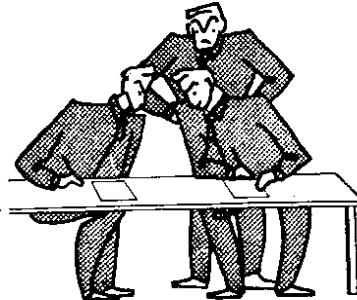
His duties include . . .



Insuring that each man has all the necessary materials



Insuring that each man understands directions



Maintaining order and preventing cheating



Excusing men who are too ill to be tested

Figure 12.

(2) In setting the stage for the test, the examiner should start with a brief informal statement explaining the test to be given, how the results will be used, and why it is important for each person to do his best on it. The aim of these remarks is a difficult one: to dispel anxiety and release tension, yet at the same time, to stress the necessity for maximal effort and output. A careless presentation may create the impression in the minds of the men that the test they are about to take is of no consequence and in no way related to their future Army careers. Or it may so impress them with the seriousness of the situation as to give rise to disturbing tensions. On the whole the best results will be achieved through a brief, straightforward but nontechnical statement of facts, delivered in a manner that is neither formidable nor severe, nor yet so jocular or perfunctory that it defeats its own purpose.

(3) Having set the stage and gained the necessary cooperation, and having distributed the test material, the examiner next informs the men exactly what they are to do. There is only one way to do this—by reading aloud the directions provided

in the manual. This is the right way, the Army way, and the only way. And it means reading aloud all the directions that are to be read aloud and no more. They should be read in a natural voice, in a smooth, coherent fashion. Hence, they must have been thoroughly practiced. Notice, however, that they should be read, not paraphrased, given from notes or memory, or adapted to someone's idea of what is more appropriate for local conditions. Every test was "zeroed-in" with its directions; the sight-setting must not be altered or the results will be wide of the mark.

(4) Nearly all Army tests are given with certain time limits which must be strictly observed if testing conditions are to be uniform from session to session, and from place to place. (See par. 26.) These time limits, either for the complete test, or for various parts of the test separately, are always specified in the test manual. They are exact, not approximate, so timing should be handled with care. Time limits should be explained carefully to the examinee. If a stop watch is available, it should be used. If not, any good watch with a second hand will serve, if used in the following manner:

(a) When giving the signal to start the test, note down on paper the hour, minute, and second of starting.

(b) Write below this time the hours, minutes, and seconds of working time for the test as specified in the manual.

(c) Add to these two figures to obtain the exact time when the signal to stop work should be given. (If the minutes add up to more than 60, of course, the sixty minutes would be carried as an additional hour and the excess listed as minutes.) Example:

Starting Time 1451:00

Time Limit for Test 45:00

---

Stopping Time 1496:00 or 1536:00

The signal to stop should, therefore, be given promptly at 1536. The timing should always be done in this way. It is unwise to trust to memory to or attempt the necessary computations mentally. And it is good practice to have some of the proctors check the timing independently.

c. COLLECTION AND DISPOSITION OF TEST MATERIALS. After the signal to stop work has been given, the materials should be collected as quickly as possible. The period between tests or at the end of a session can be one of tremendous confusion with everyone talking, comparing notes, reaching for coats and hats, and being anxious to leave. Under these circumstances the test materials are apt to be collected in haphazard fashion, with booklets and answer sheets jumbled together. And in the confusion booklets have a way of disappearing. Remember that Army tests are classified as restricted and must be accounted for. Remember also that the booklets will be used again and the answer sheets have to be scored. Planned and orderly collection of materials will pay big dividends in the time saved during these future operations. The following system achieves a maximum of order and control.

(1) As soon as the stop signal is given, the examiner should instruct the men to remain quietly in their seats and to follow directions in order to expedite the collection of materials.

(2) He should then direct them to pass these materials to the ends of the rows, specifying which end. The materials should be passed separately, first answer sheets, then test booklets, then supplementary materials, such as scratch paper, and finally, pencils.

(3) The men at the ends of the rows should be instructed to stack the materials in separate piles, making sure that all booklets are closed, with the

cover sheet outside, and that all answer sheets are faced the same way.

(4) When this is done, the proctors can collect the materials and at the same time make a rapid count of the numbers turned in. Only after all materials have been checked in and accounted for should the group be dismissed or the next test begun.

d. CARE OF BOOKLETS. After each testing session, and certainly before the next session, all test booklets should be carefully scrutinized for answers or marks of any kind. In spite of all warnings, some persons will write answers in the booklets or use them for scratch paper. If the answers or marks can be erased, this should be done. If not, or if the booklet is worn or torn, it should be destroyed in accordance with paragraph 60, AR 380-5. All used scratch paper should be destroyed. All tests and testing supplies should be kept under lock and key when not in use.

### Section III. DIRECTIONS FOR SCORING GROUP TESTS

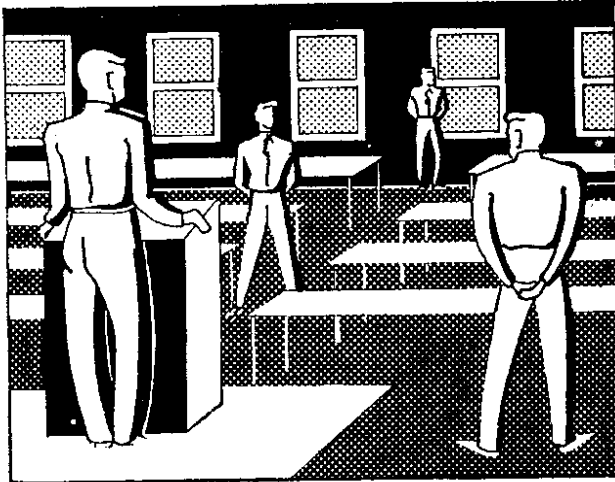
#### 46. General

Scoring a test is fundamentally the procedure by which the number of correct answers is counted. In many instances this simple process is elaborated. Some tests of the multiple-choice answer type, for reasons discussed in the previous chapter, are scored by counting the number of correct answers, and subtracting from this figure some fraction of the number of wrong answers. If this is to be done, the manual will so state, and will indicate what proportion of the wrong answers is to be subtracted. The statement "right minus one-fourth wrong," for example, means that one-fourth of the number of wrong answers is to be deducted from the number of right answers. This statement is known as the *scoring formula* for the test. One must be sure to use the proper scoring formula for each test as given in the manual for that test and to follow it exactly. (See par. 30a.)

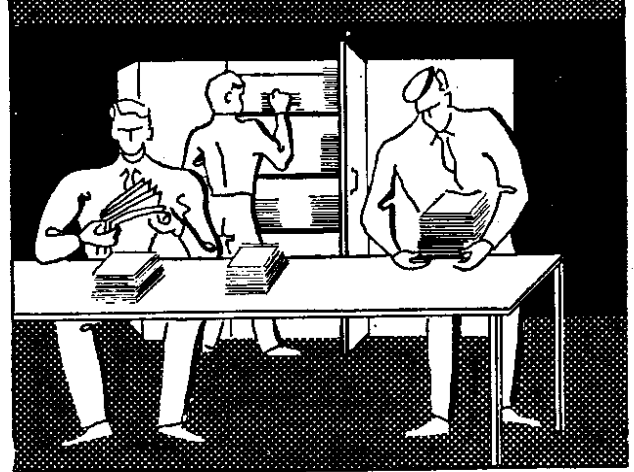
#### 47. Hand Scoring and Machine Scoring

Some tests must be hand scored, and others, using special answer sheets and special pencils, may be scored by means of the International Test Scoring Machine. Both methods may make use of a scoring formula, but with the scoring machine the correction can be made automatically. However, some tests are beyond the capacities of the machines. Which

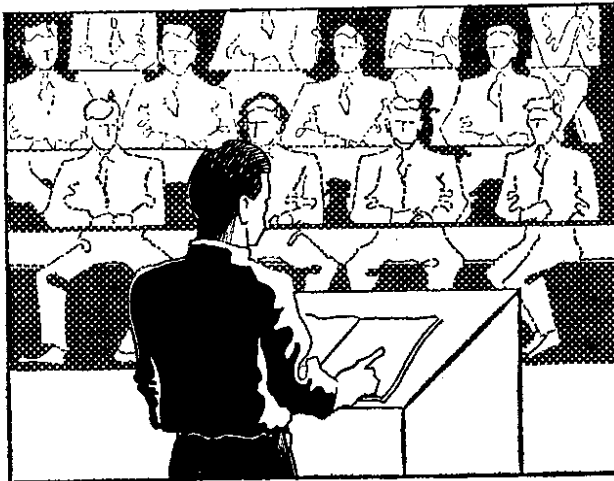
# STEPS IN THE ADMINISTRATION OF GROUP TESTS



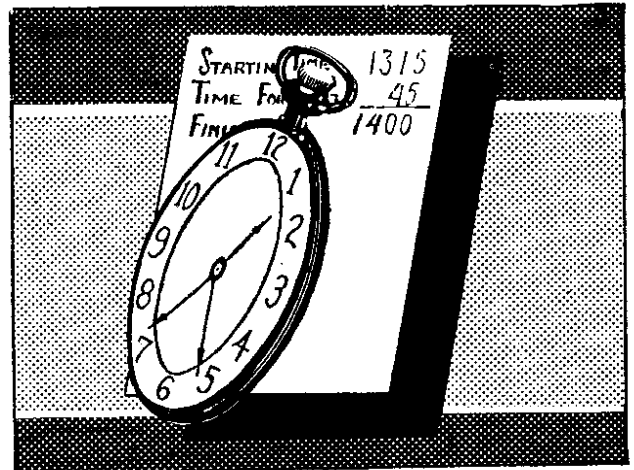
1. REHEARSE THOROUGHLY



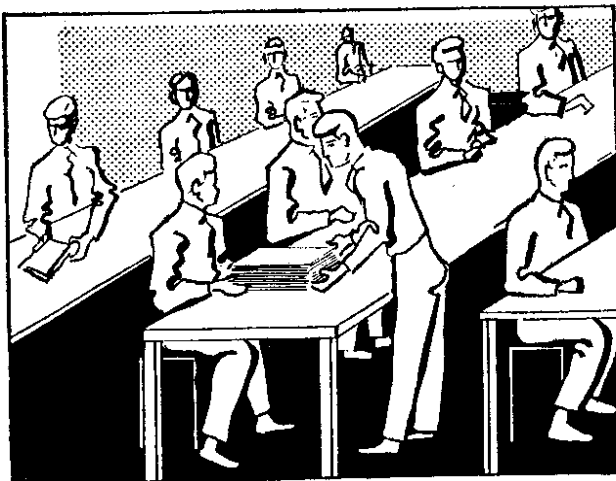
2. PREPARE MATERIAL IN ADVANCE



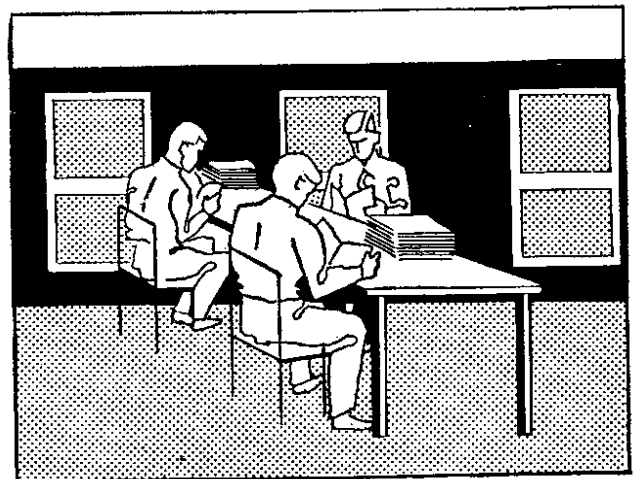
3. FOLLOW THE MANUAL EXACTLY



4. TIME THE TEST ACCURATELY



5. COLLECT MATERIALS CAREFULLY



6. CHECK TEST BOOKLETS FOR MARKINGS

Figure 13.



method is employed will, therefore, depend upon the nature of the test and the availability of the machine. In this connection, it is important to note that all machine-scorable tests may also be hand scored.

#### 48. Scoring Team

a. Whether scoring is done by hand or by machine, there are a number of steps or operations that must be performed in orderly sequence. Efficient scoring will require a team, with a different individual assigned to each of the successive steps in the process. Members of the team will complete only the operations to which they are assigned, passing the papers along to others who perform succeeding operations. With hand-scored tests of the type having the answers on the test page itself, it is efficient to have each scorer handle a single page, counting the number of correct answers and writing this at the bottom of the page. Other members of the team should be designated to add together the scores for each page and convert total scored into standard score terms; still others should check each of the steps in the process. A convenient method of handling machine-scored answer sheets requires a team composed of one man to do each of the following (the operations are described in later paragraphs):

- (1) Scan answer sheets.
- (2) Count attempts.
- (3) Convert raw scores to standard scores.
- (4) Operate the scoring machine.
- (5) Check the conversion.
- (6) Spot check (every 25th answer sheet, for example) by hand scoring.

b. The importance of checking throughout cannot be overemphasized. If the test is worth enough to be allotted an hour of the examinee's time, it is certainly worth the additional minute or two required to insure that the score is an accurate one.

#### 49. Instructions for Hand Scoring

Hand-scored tests are of two types: expendable tests in which the answers are made directly on the booklet itself, and nonexpendable tests with which a separate answer sheet is used. With tests of the first type, scoring keys, or scoring stencils, or merely a list of correct answers may be provided. It is advisable to devise some form of key or stencil if none is provided. With a few Army tests, such as the Qualification Test (Q-1 or -2) and the Classifi-

cation Test R1, the answers will be check marks or figures or underlined words scattered over the face of the test itself. With tests of this kind, the stencil is most efficient. Satisfactory stencils can be made with large sheets of clear celluloid or plastic (exposed X-ray films, for instance) cut away in the places where the correct answers will appear when the stencil is laid over the test. In this way the answers appearing in the windows of the stencil can be checked and counted with ease. The transparency of the stencil adds the advantage that answers on other parts of the paper, wrong answers, can also be counted without removing the stencil. Other expendable tests are arranged so that the answers—check marks, underlinings, write-ins, etc.—are entered in a column running down the right-hand edge of the test page. Scoring such tests is facilitated by the use of a strip key—a strip of cardboard with the correct answers indicated with proper spacing so that when set down alongside the column of answers on the test page, the examinee's answers and the correct answers can be directly compared. Both stencils and strip keys can be readily improvised in the field.

#### 50. Hand Scoring Separate Answer Sheets

Many Army tests make use of separate answer sheets on which the answers are indicated by a black line in the proper space. With tests of this type, scoring keys are usually provided. These consist of opaque cards that fit over the answer sheet and have holes punched in the positions in which the correct marks will appear. In using a key of this kind, provision should always be made for aligning the key with the edges of the paper, or preferably with two fixed "landmarks" on opposite corners of the answer sheet. Much time in lining up the key will be saved thereby, and increased accuracy will be gained. The procedures outlined below for scoring papers and for providing all necessary checks have proved valuable over a period of years and are strongly recommended. In using this method, one will require, besides the punched-hole scoring key, two pencils of different color, red and blue, for example. The method is essentially one of counting right answers, wrong answers, and omissions (questions with two or more answers marked are counted as omissions). The successive steps in the method are:

- a. Look over each answer sheet and draw a red pencil mark horizontally through all response positions for each item for which the examinee has made

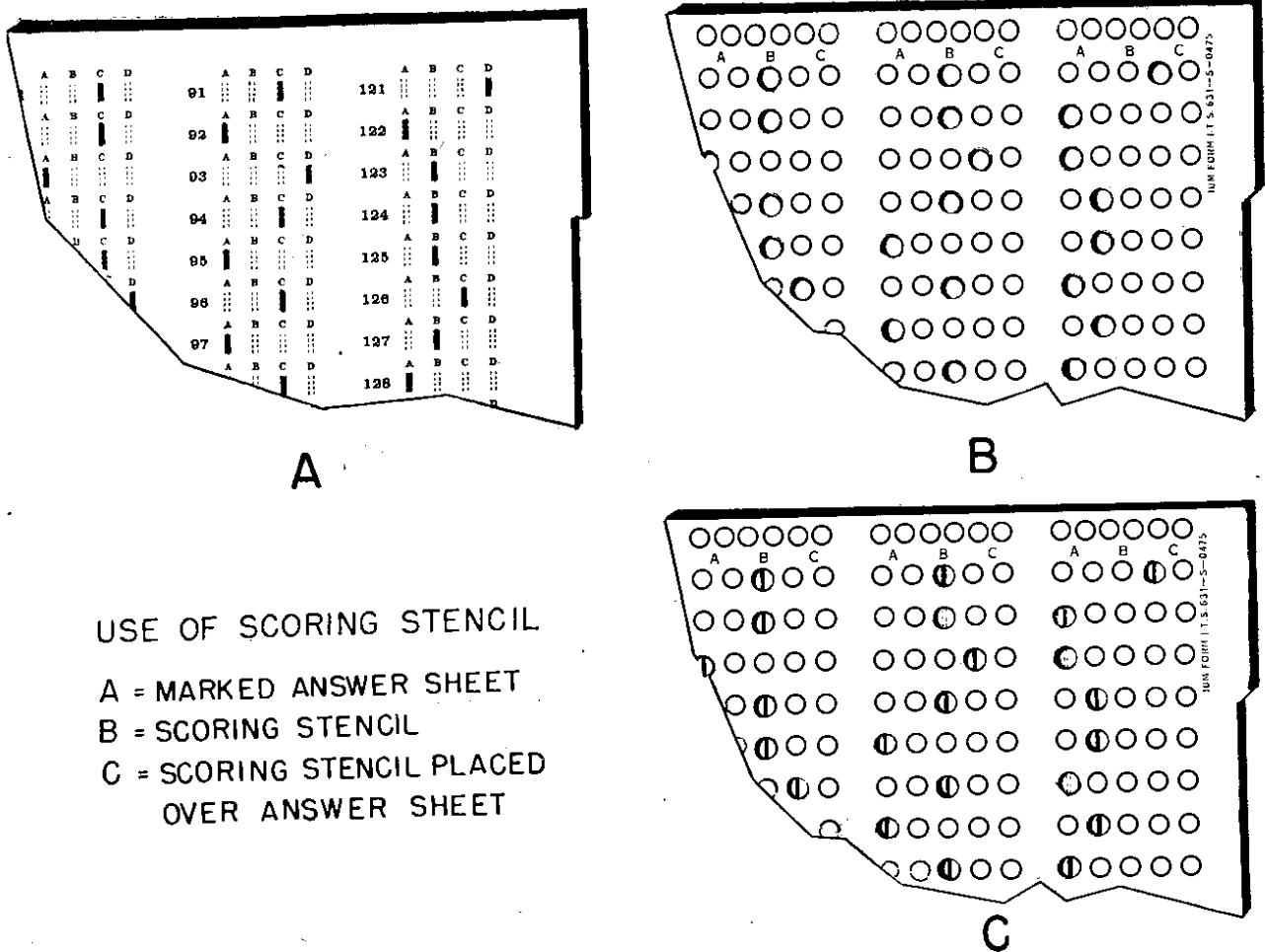


Figure 14.

more than one choice or no choice. Care should be taken not to mark in this way reponse positions for which the intended choice is clearly indicated even though more than one is marked. The sum of these red marks will be the number of *omissions*.

b. Place the punched-hole key, right side up, over the answer sheet and register it. Count the number of black marks made by the examinee, which appear through the holes of the key, excluding, of course, those black marks with a red line running through them (multiple answers). The sum of these black marks will be the number of *right* answers.

c. With the key still in place, draw an X with the blue pencil on the answer sheet through each hole where there appears neither a black mark (right answer) nor a horizontal red mark (omission). The sum of these blue marks will be the number of *wrong* answers.

d. If the scoring formula specifies that the raw score is merely the number of right answers, only steps *a* and *b* need be performed. If some fraction of the number of wrong answers is to be deducted from the number of right answers, the figures obtained from steps *b* and *c* will be entered in the proper formula to obtain the raw score.

e. A check is provided, in that the number of omissions (step *a*) plus the number of right answers (step *b*) plus the number of wrong answers (step *c*) should equal the total number of items on the test (exclusive of practice items).

f. There are two additional sources of error that must be guarded against. If the test does not utilize all of the response spaces on the answer sheet, care must be exercised to insure that no marks beyond the last question are counted. In other words, the sum of omissions, right answers, and wrong answers must not be more than the number

of items on the test. The same precautions must be taken to avoid counting any of the practice items.

### 51. Machine Scoring Answer Sheets

The widespread use of tests in the Army classification system would have been wellnigh impossible without the scoring machine. With this delicate and intricate instrument, scoring which would require the labor of several men for hours can be handled by one man in a much shorter time. The principle of the International Test Scoring Machine is simple. The graphite deposited by a special lead pencil in making a mark on paper will conduct an electric current. If two wires from a source of power are pressed against such a mark, the circuit will be completed. The current will be carried from one wire through the mark to the other wire and it will cause a deflection of the needle of a galvanometer connected in series in the current. If there are hundreds of these simple circuits, all connected to the same galvanometer, all of those which are closed by means of pencil marks will add to the current flowing through the galvanometer. In other words, the amount of the deflection will tell how many of the circuits are completed. In a sense, therefore, the galvanometer reading is a count of the number of pencil marks. If the answers to a test are indicated by soft lead pencil marks in a specified place on an answer sheet, and if this answer sheet is then pressed up against a mass of open-end circuits (or electrodes) the dial of the galvanometer will register the number of such marks. But if a punched-hole scoring key is inserted between the answer sheet and the electrodes, the current carried by the "right" pencil marks can be routed one way, and the current carried by the remaining marks routed another way. Thus the dial can be made to register the number of right answers, the number of wrong answers, the number right plus the number wrong, and finally the number right minus any portion of the number wrong—all at the will of the operator and the turn of a switch. This general account is provided to furnish understanding of the machine requisite to accurate scoring and to help the operator guard against its limitations.\*

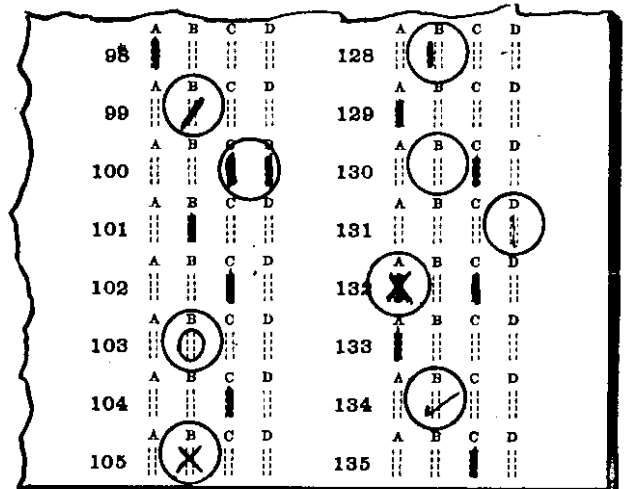
a. The scoring machine is an amazingly intricate and wonderful device, but is not human. While

\*A technical description of the International Test Scoring Machine will be found in the Manual of Instruction for the International Test Scoring Machine, C. R. 9145 (revised).

this is something of an advantage, insofar as it eliminates some of the human frailties of subjectivity and inaccuracy, still it limits the machine's powers of discrimination. It cannot, for example, tell the difference between an intended answer and a stray pencil mark, and will count both indiscriminantly. Also, it cannot count a pencil mark if this is not brought in contact with the electrodes. For these reasons a specially printed answer sheet, with response spaces properly located, must be employed. Since not all pencil leads contain the necessary ingredients, a special pencil must be used and a good solid mark must be made to indicate answers.

b. Because the scoring machine is, after all, a mechanical device, the answer sheets must be carefully prepared for scoring if accurate results are to be obtained. This procedure is called *scanning*, and consists of a thorough check of each of the following points. Make sure:

- (1) That each pencil mark is heavy and black. Light marks should be gone over with the special pencil.
- (2) That each mark is in the space between the pair of dotted lines and entirely fills this space.
- (3) That all stray pencil marks on the paper clearly not intended as answers are erased.
- (4) That no response position for any question has more than one answer indicated. If such multiple answers occur, all marks in the response position should be thoroughly erased, and that item



**WATCH FOR THESE ERRORS WHEN  
SCANNING ANSWER SHEETS**

Figure 15.

considered as an omission. The answer sheet should be hand scored if erasures are impractical.

c. The operator must familiarize himself with the method of inserting the keys, of checking and adjusting the scoring circuits, and of using the formula switches. All of this material is contained in the Manual of Instruction for the International Test Scoring Machine which accompanies each instrument. For utmost accuracy with the use of the machine, the following method of scoring is strongly recommended. After the keys have been inserted, the scoring circuits checked and adjusted, and the answer sheets scanned (see *b* above), the successive steps of the method are:

(1) Count the number of questions attempted and enter this figure in one of the score spaces at the edge of the answer sheet. It will usually be simpler to count the omissions (remembering that multiple answers are erased and counted as omissions) and subtract this figure from the total number of questions.

(2) Insert the answer sheet into the machine and set the formula switch at R + W (right plus wrong). The resultant meter reading should be the same as the number of questions attempted [step (1)]. If these two figures differ by more than one point, the paper should be set aside for further scanning or for hand scoring.

(3) If the two values check, set the formula switch according to the desired scoring formula. The resultant reading is the raw score for the test and should be recorded in a designated score space at the upper edge of the answer sheet.

d. Accumulated experience has demonstrated that the scoring machine, properly used, will result in greater accuracy than the average hand-scoring methods. To insure the maintenance of this high level of accuracy, special attention is directed to the following precautions:

(1) At the beginning of each day's scoring, and after each 100 answer sheets have been scored, check the scoring circuits according to the methods outlined in the Manual of Instruction for the International Test Scoring Machine. If the machine is not working properly, get in touch with the nearest IBM representative, and use hand-scoring methods until it has been satisfactorily adjusted.

(2) Rescore by hand-scoring methods a random sample of answer sheets each day as an additional check on scoring procedure.

(3) On damp days use the heating unit to dry the papers preparatory to scoring. Moisture also is a conductor of electricity.

## Section IV. DIRECTIONS FOR ADMINISTERING INDIVIDUAL TESTS

### 52. General

For the most part, the tests employed by the Army classification system are of the paper-and-pencil type administered to groups of examinees at one time. In certain special circumstances, proper classification and disposition of the soldier will demand the administration of an individual test. The expenditure of time and effort is greatly increased with the employment of such instruments, and their use must be limited. In general, the individual type of test is recommended for cases in the following categories:

a. Where the paper-and-pencil type test is inappropriate because the subject is lacking in the educational skills of reading and writing.

b. Where more personal contact is needed to insure that the examinee is at ease, is properly motivated and encouraged, and knows just what he is supposed to do.

c. Where it is desired not only to determine the individual's over-all score, but also to give the examiner the opportunity to observe him at work and to estimate his specific strengths and weaknesses.

### 53. Individual Testing Session

The individual testing session is a more personal and somewhat less formal affair than the group testing session. This does not make it easier to manage; on the contrary, the individual test administrator needs much more than average training and experience to achieve results which can be accepted with any confidence. His task is a difficult one. He must administer a rigidly controlled and carefully standardized set of problems under conditions precisely as specified, all the while creating an impression of friendly informality. Men assigned to this job must be selected with care. The ideal examiner is a man with a knowledge of the principles of psychological measurement and an appreciation of the needs for exactness and precision. He is personable and friendly and an easy conversationalist. He is patient and tolerant, never given to a show of arrogance, flippancy, or sarcasm, no matter how absurd the responses of the subject might be. And he is a monument of tactfulness.

a. Individual tests, because they are essentially personal interviews, should be given in an atmosphere of privacy. Special rooms are recommended, but separate booths divided by partitions

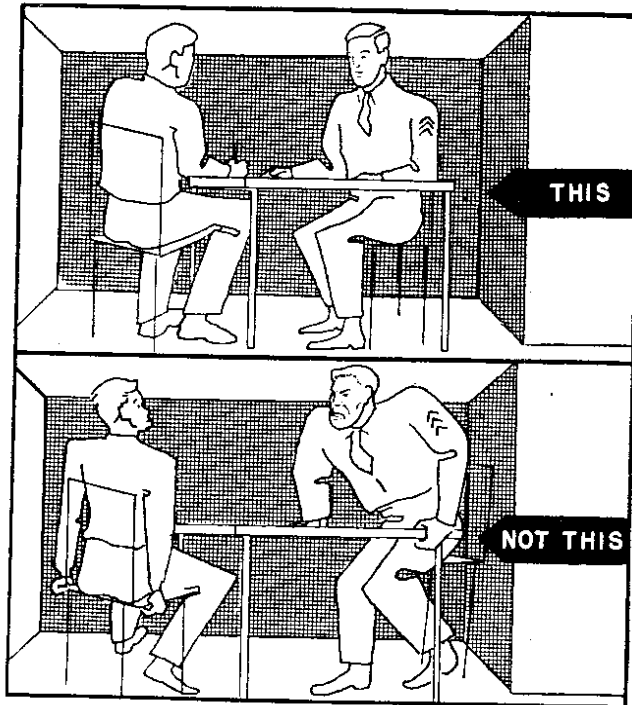


Figure 16.

will serve where facilities are limited. Examiner and subject should be provided with straight-back chairs of a comfortable height, facing each other across a table or desk. This table or other working surface should be large enough to accommodate all of the test materials and provided room for the examiner to jot down responses on the record sheet. The common field table is of the proper dimensions for most individual testing.

b. Before undertaking the administration of an individual test, the examiner should make a careful study of the manual of directions and of the test materials. The exact wording to be used in presenting the materials will be specified, and should be reshearsed until it can be read in a normal conversational manner. The examiner should also practice the things he is to do—the placement of the materials, movements, pointings, demonstrations, etc.—until these are smoothly coordinated with the verbal directions. Finally, he should give practice administrations with “guinea pig” subjects, under the supervision of a qualified examiner until he can maintain the subject’s interest and confidence, select and use necessary materials and instructions without fumbling, and make of the whole procedure a smooth and effective performance.

c. The administration of an individual test begins as soon as the examinee enters the room.

The first step is to get him into the proper mental condition for taking the test; to remove fear and tension which may conceal qualities valuable to the Army. The man reporting for the test is quite apt to be afraid, discouraged, misinformed, or truculent, and in no condition to perform in a fashion that can be considered a trustworthy indication of his true ability. Here the care exercised in selecting examiners will begin to pay dividends. The skillful examiner will greet the subject in an affable manner, ask him questions about himself, about his work, listen to his complaints, and give every indication of being truly interested. He will often have to call upon all his skill and patience to carry this off without creating an air of forced and stilted play acting.

d. The transition from this informal chat to the presentation of the test materials should be gradual and natural. The test manual will probably contain suggestions for bridging this gap, or such statements as “I have some problems here I would like you to try,” or “Let’s see what you can do with these questions,” will serve the purpose. From this point on, the presentation of the problems, the questions, and all directions to the subject must follow the exact wording of the manual. Moreover, any performance materials, such as blocks or pictures, must be placed on the table or exposed precisely as specified. And the different parts of the test must be given in order, without skipping around. It cannot be overstressed that any departure from the manner of administration in which the test was constructed and standardized will make the test score unreliable.

e. The examiner should speak distinctly and slowly while administering the test so that the examinee may hear and understand, for the examiner may not repeat any questions (unless some unexpected disturbance has prevented his being heard the first time). He must guard against gestures, words, or inflections of the voice that may suggest an answer. Throughout the course of the examination, the examiner will have to stimulate the subject to do his best. He will have to make appropriate remarks of approval or praise after each success, and he will have to console and encourage him when he fails. Suggestions for such appropriate remarks will usually be included in the manual for administering the test.

## 54. Timing

With the individual test, the problem of timing is usually somewhat more difficult than with group

tests. This is so because the problems and questions which compose the individual type of test often have separate time limits, and these limits are often specified in terms of seconds rather than large intervals. Furthermore, many of the items of an individual test are scored in terms of the time required to arrive at the correct solution. This means that the examiner will have to look at his watch frequently. Yet, because of the close personal nature of the situation, the examiner cannot be too obviously engrossed in the time problem without creating a disturbing and distracting tension in the examinee. Everyone has experienced the feeling of nervous strain when working against time and the maddening tendency of fingers to become all thumbs as the seconds tick off. Some of this tension is natural, of course, when one is told to work as rapidly as he can; but it is tremendously heightened if the examiner is a nervous clock watcher. So, the timing should be done unobtrusively and with the appearance of casualness. This does not mean, however, that it can be slipshod. It is of utmost importance that the timing be precise and that the exact limits specified in the manual be observed. Where tests are scored in terms of time, an error of a few seconds may account for a difference of several points of score. It is essential that the examiner should be thoroughly practiced in timing. If possible, he should use a stop watch, because this will always start at zero, and because the timing can be done with one hand, leaving the attention of the examiner focused on the test itself. If a stop watch is unobtainable, an ordinary watch will be used. It should be of the type equipped with a large sweep second hand for ease of reading, and the examiner should always give the starting signal for any item when the second hand is at zero.

## Section V. SCORING INDIVIDUAL TESTS

### 55. General

In addition to administering and timing the parts of the test, the examiner must record and score the subject's responses. In doing this, he will be concerned with a record or score sheet and with lists of scoring criteria and time credits contained in the manual for the test. He should study these carefully and use them precisely as directed.

### 56. Score Sheet

The score sheet for any particular individual test is usually a single page upon which are arranged the

spaces for recording or indicating answers and scores. On this page will be separate areas for each of the subparts composing the test, and these in turn will be lined to accommodate the answers to individual items. The record of the response to a particular item may be a mark (plus or minus), a time interval, or a word or phrase. The score sheet will usually contain all scoring credits and weights to facilitate scoring. For instance, if each correct response is to be allowed a certain number of credits, these can be read off directly, entered in another column, and summed to yield the total score for that subpart. Finally, the record sheet will usually provide a section for the recapitulation of part scores, for any multiplying factors employed, and for entering subscores, total scores, and converted scores. The examiner jots down the marks, words, phrases, or times in the appropriate rows and columns as the testing session progresses, and, where possible, he allots to each response the credit it has earned at the time it is given. At the end of the examination, he can verify any doubtful answers by reference to the scoring criteria in the manual, add the credits, and compute the various scores. In doing all of this, he should avoid directing the examinee's attention to the machinery of scoring, or to any particular score credit given. It is advisable to hold the manual with the left hand in such a way that it shields the score sheet from the subject's view.

### 57. Scoring Criteria

The scoring criteria and tables of time credits will be included in the manual for the test. The nature of the individual testing session makes the subject's responses less circumscribed than responses to the paper-and-pencil type test. The individual test examinee will seldom be required, for example, to choose one of four possible answers or to respond with a single word or figure. More often he will be asked to explain or discuss in his own words or to offer his solution to a problem. The fact that the response is freer should not, however, imply that it is less objective. In order to provide the desired objectivity and uniformity in scoring, it is necessary to draw up certain criteria to be used as standards. These criteria will be in the nature of lists of all possible acceptable answers or of examples of the type of answer for which full or part credit will be allowed. They will also sometimes list nonacceptable answers for which no credit is to be given. Faithful, even slavish, adherence to these criteria

**ARMY INDIVIDUAL TEST**  
Score Sheet, AIT-I, S

Name \_\_\_\_\_ Age \_\_\_\_\_ Date \_\_\_\_\_  
Last First Middle Initial Years

Army Serial Number \_\_\_\_\_ Organization \_\_\_\_\_

**SUMMARY RECORD**

Test	Verbal Tests	Test Raw Score	Weighted Raw Score
I.	Story Memory	x1	
II.	Similarities-Differences	x2	
III.	Digit Span	x4	
Total: Verbal Raw Score			

Test	Performance Tests	Test Raw Score	Weighted Raw Score
IV.	Shoulder Patch	x1	
V.	Trail Making	x2	
VI.	Block Assembly	x1	
Total: Performance Raw Score			

FD AGO FMT-20

**PERFORMANCE TESTS**  
TEST IV. SHOULDER PATCH

Test	Order of Cards	Time (Seconds)	Accuracy	Time	Total (A+B)
1.	1-12				
2.	5-11-16				
3.	3-13-12-17				
4.	5-15-14-13				
5.	4-6-10-12**				
6.	2-11-16-17-9				
7.	3-15-17-16-14				
8.	5-11-13-14-7				
9.	5-12-6-7				
Test Raw Score					

\*\*also 13-9-10-17 or 11-6-10-17

**TEST V. TRAIL MAKING**

Part	Accuracy	Time	Score
A.			
B.			
Test Raw Score			

**TEST VI. BLOCK ASSEMBLY**

Part	Accuracy	Time	Score
A. 5-block			
B. 12-block			
C. 24-block			
Test Raw Score			

Figure 17. Example of an individual test score sheet.

is a primary requisite for uniform and accurate scoring. The examiner should adopt the strict rule that he will never give credit for a response not indicated as acceptable no matter how satisfactory or plausible it may seem to him. If the response in question is not among those listed, one may safely assume that it has been omitted—not through oversight—but because it is not a wholly accurate response or one typical of the best solutions to the problem. Again it will be helpful to remember that tests are standardized prior to release according to the directions and criteria contained in the manual, and any deviation from these can only result in a loss of uniformity. Occasionally, it is true, a rare but correct answer will crop up. But before this can be credited as correct, it should be analyzed and included, officially, in a revision of the criteria. A common weakness of examiners is the tendency to allot credit for what they think the

subjects mean to say rather than for what they do say. The urge to do this may spring from a desire to help out, or may result from a feeling that the examinee knows more than his answer would imply. In either event the tendency should be avoided. If examiners were unerring judges of human capacities, testing would not be necessary, but experience has abundantly demonstrated that tests are far more reliable than individual "hunches."

## Section VI. OBTAINING STANDARD SCORES

### 58. General

The number of correct answers to a test, or the number of right answers less some portion of the number wrong, or in the case of an individual test, the sum of all the individual credits allowed is called the *raw score*. For reasons that will be made clear

in the next chapter, this raw score is seldom the one which is reported or made a matter of record. With most Army tests a uniform scale of standard scores is employed, and the last step in the examining and scoring procedure involves the conversion of the obtained raw score into standard score form. In practice this is a simple matter of using the proper conversion table which will be found in the manual for the test. These tables are drawn up in two columns, the first for all possible raw scores that can be obtained on the test, arranged in order of descending magnitude, and the second for the corresponding standard scores. The total range of scores is generally also divided into five Army grade intervals. Having arrived at a certain raw score for the test, the examiner or scorer can locate this in the first column of the conversion table and read across to the corresponding standard score. Then,

by reference to the grade divisions, he can determine the Army grade into which it falls. As in all purely clerical activities, however, errors frequently occur. Therefore, the conversion should always be given an independent check.

### 59. Summary

The standard score and Army grade are the end results of the administering and scoring process. If the principles and procedures set forth in this chapter are faithfully followed, scores will be obtained that can be accepted with confidence as true measures of the examinees' skills and capacities. The manner in which these results are to be interpreted and the uses to which they are put in the general system of classification in the Army—these will be the subjects of later chapters.



## CHAPTER 5

## THE MEANING AND INTERPRETATION OF TEST SCORES

## Section I. THE ARMY STANDARD SCORE SCALE

## 60. Review and Orientation

Scores on most Army tests are reported and recorded in the form of *standard scores*. As has been stated (par. 35) this is done because such standard scores make analysis of test data more efficient; they reveal in a way that is clear and easy to grasp the information required by personnel concerned with classification. Each particular Army standard score is, in effect, a summary not only of the data concerning the performance of a man on a particular test, but also a summary of the significance of his performance. A man's test performance has the following general kinds of significance to the Army: it tells how the man compares with others as to the amount of a particular aptitude or skill he possesses; and, coupled with a statement of the test's validity, it tells approximately what performance on a particular assignment may be expected of a man, and how the performance expected of him compares with that expected of others.

## 61. Development of Standard Score Scale

The most practicable methods of summarizing test data and their significance are statistical, and the standard score scale is, therefore, a statistical development. For all Army tests, the necessary data are computed by expert technicians and supplied in the form of convenient conversion tables. However, the concept of the standard score will be interpreted more correctly and used more effectively if all classification personnel gain a general understanding of the statistical techniques involved.

a. OTHER TYPES OF SCORES. A brief analysis of the limitations of other scoring systems will make clear many of the advantages of the standard score scale.

(1) The familiar system which attaches arbitrary values to percentages, as for example 60 percent meaning "poor," 80 percent meaning "good," 90 percent meaning "excellent," are not useful to the Army. Such scores, when achieved on different tests, cannot be compared. Ninety percent on an arithmetic test may represent an entirely different degree of ability than 90 percent on a lan-

guage test. Arbitrary percentage scores, therefore, do not indicate whether a man may be better at mathematics or languages. Moreover, arbitrary values given to percentages often depend upon the judgment of the examiner.

(2) Test results in the form of raw scores lack meaning. The number of questions which an examinee answers correctly is partly a function of the number of items in the test. And since it is impossible, for practical reasons, to make all tests uniform in length, or to attach uniform weight to all test items, raw scores on two different tests cannot be compared directly. A raw score of 50 on a reading aptitude test, for example, may be as different from a raw score of 50 on a mechanical aptitude test as is 50 acres from 50 war bonds. *It is necessary to translate raw scores into values which are comparable.* It is necessary, in other words, to find a "medium of exchange" for the field of mental measurement. Such a medium would make it possible to compare all measurable characteristics, just as money makes it possible to compare the wealth of the man who has 50 war bonds with that of the man who has 50 acres.

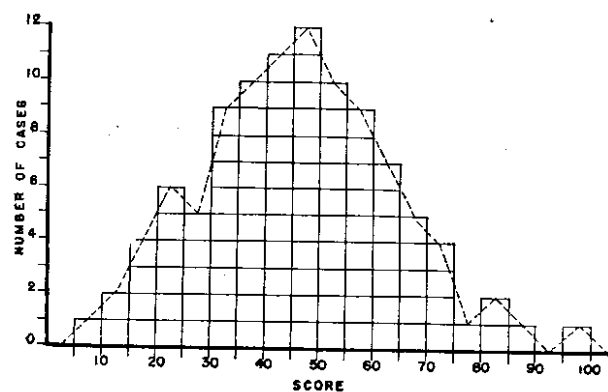


Figure 18. Distribution of test scores. The figures along the base line indicate the range of scores from lowest to highest. Each block represents a man and is placed above the score (or score interval) indicating his performance on the test. Thus, two men scored in the interval 10-14, four in the interval 15-19, etc. A line drawn to connect the tops of each of these columns of scores shows the general shape of the distribution.

(a) If the scores made by all persons of a group are plotted, the resulting figure is called a *score dis-*

tribution or a distribution of scores. The method of plotting is illustrated in figure 18.

(b) Most distributions of test scores plotted from data obtained on large unselected samples approximate that shown in figure 19. This curve is symmetrical and bell-shaped. A vertical line drawn through its highest point divides its area into two equal parts; and the point where this line intersects the base line is the *mean* or average score of the distribution. This curve, called the *normal curve of distribution*, is actually a mathematical abstraction. It is, however, sufficiently typical of most curves representing Army test scores to be useful as a basis of comparison. Moreover, the percentage of men who stand higher or lower than any given point on a normal curve closely approximates the proportion above and below the same point on distribution curves of Army test scores. The normal curve thus has great usefulness in analysis and prediction.

(c) Several characteristics common to every distribution of scores should be noted. The width of the curve along the base line indicates the *range* of scores in the group tested. The height of the curve at any point above the base line indicates the number of persons receiving that particular score. The area bounded by the curve and the base line represents the number of cases (*N*) in the group tested.

(d) The shape which a distribution of raw test scores will assume depends upon several characteristics of the test and the group tested. If a test contains a large number of items, the distribution of its raw scores will extend over a wider range than will be the case for a shorter test. Moreover, depending upon the number of items and their difficulty, the position of the average score will vary for different tests. Thus, a particular raw score may be better-than-average for one test and below the average for another. Under these circumstances it is obvious that *raw scores* are not comparable or meaningful.

b. ARMY SCORES. (1) All Army scores are given the same average and the same *spread* or *range*. The distribution curves, therefore, have the same shape; and points in the same relative positions (above or below average) have the same score values. Army scores are, therefore, comparable and their meaning in terms of relative ability is made clear. Distribution of scores for each test are given the same range and reference point by translating the raw scores into scores on a standard scale. This scale is developed by selecting the av-

erage of the distribution as a base point, and expressing the distance (or deviation) of each score from that point in terms of the deviations of all other scores. A mathematical expression of the deviations of all scores from the average of a distribution is found as follows:

(a) Subtract the average score from each raw score.

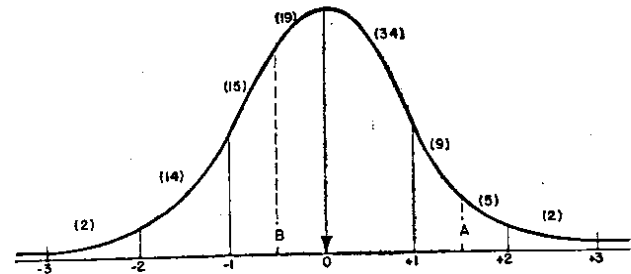


Figure 19. A normal distribution with average (shown by arrow) and standard deviations indicated. A score at A, regardless of its size in raw score terms, has a standard score value of +1.5 since it is half way between one and two standard deviation units ABOVE the average. Likewise, B represents a standard score of -0.5 since it is half of a standard deviation unit BELOW the average. Each man's standard score indicates his relative standing in the group, since the mathematics of the normal curve allow determination of the percentage of the total area bounded by any two standard scores. The numbers in parentheses show some of these percentages. It can be seen that the score at A is exceeded by 7% of all scores in the group, and that the score at B is higher than the scores of 31% (15 + 14 + 2). Furthermore, 62% of the group score between A and B (19 + 34 + 9). Interpretations in these terms are meaningful.

(b) Square each of the remainders, or deviations, so obtained.

(c) Add all of the squared deviations together.

(d) Divide this sum by the total number of scores.

(e) Extract the square root of this quotient.

(2) The quantity thus obtained is called the *standard deviation* of the distribution. It is, in effect, a standard unit with which the deviations of all individual scores may be compared. Thus, any raw score can be converted into a standard score by computing its difference from the average raw score and dividing by the standard deviation of the raw scores. Figure 19 shows a normal curve of distribution with the average and standard deviations indicated.

c. ARMY STANDARD SCORE SCALE. The standard score scale is so short—usually from about -3 to +3—that in order to make a fine enough differentiation between men, it is necessary to resort to decimal scores. Another inconvenience is that all

scores below average are negative in sign. The Army, therefore, multiplies each standard score by 20 in order to get rid of the decimals, and adds 100 to each score to get rid of the negative signs. Thus, the Army standard score scale designates the averages as 100, and each standard deviation is represented by 20 score points. Army standard scores, therefore, range from approximately 40 to 160 with about 68 percent of all cases falling between 80 and 120. (See fig. 20.)

d. **CONVERSION TABLES.** The Army provides tables with its tests so that the raw scores recorded at any time and place can be readily converted into Army standard scores. Each of these tables is based upon the distribution of scores achieved on the test by the standardization population. (See par. 36.) Thus, when converted the score of a particular individual at a particular installation is compared with a very large and representative group of Army men and each individual's rank, as regards the trait measured, is thus given in terms of the Army as a whole. All individuals are compared with the same standards. Army standard scores obtained by use of authorized conversion tables are, therefore, much more useful in classification than would be such scores derived by direct statistical computation from local data collected at any single installation. Moreover, the use of these conversion tables insures uniformity in translating raw into standard scores.

## 62. Advantages of Army Standard Scores

Army standard scores are most useful for classification and assignment purposes by reason of the following specific advantages. They state test performance in such a way that small individual differences in potential ability or achievement are clearly revealed. They tell how a man ranks in comparison with other Army men. They make it possible to compare an individual's expected performance with that of others, and to compare each man's performance on any number of tests. They are mathematically convenient and, therefore, make further statistical analysis of data more efficient.

## 63. Army Grades

a. For many purposes of military classification, the Army standard score scale is a more refined measuring device than is required. In practice, a statement of the general range of potential abilities or achievements within which a man falls is in many cases sufficient. Army grades provide such a general statement. They give a rough, handy, indication of an individual's position in the Army-wide distribution of the characteristic concerned. Figure 20 indicates the meaning of Army grades in terms of Army standard scores and also in standard deviations from the average. Note that each Army grade includes the same range of standard scores,\* but that the proportion of men is largest in Grade III, smaller in Grades IV and II, and

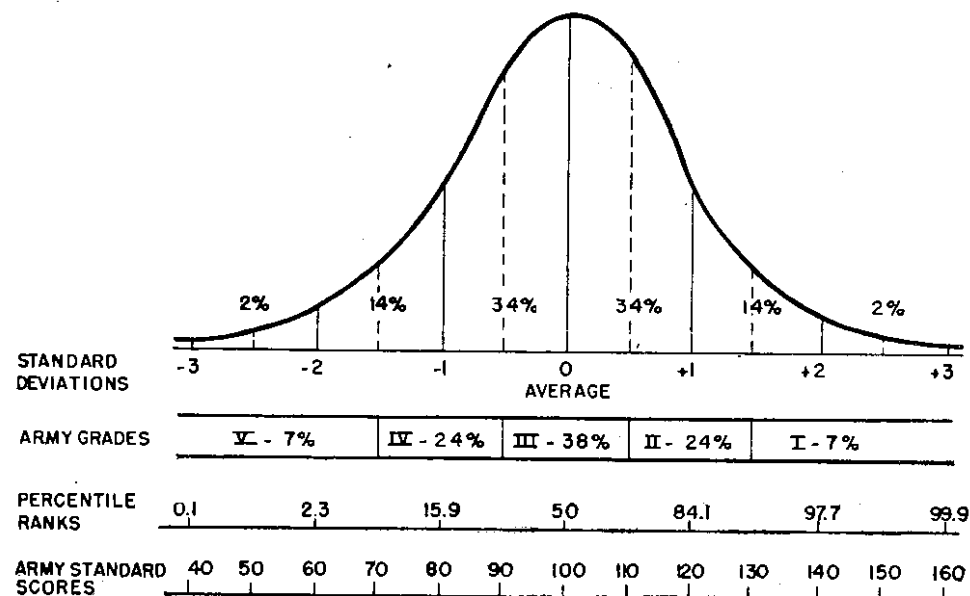


Figure 20. Army standard scores, Army grades, and percentile ranks in relation to the normal curve of distribution.

\*In order to compensate for irregularities in the distribution of scores, the limits for the Army Grades IV and V on the Army General Classification Test have been adjusted. These limits are: Grade V = 42-59; Grade IV = 60-89 (Army standard scores). This variation holds only for the AGCT.

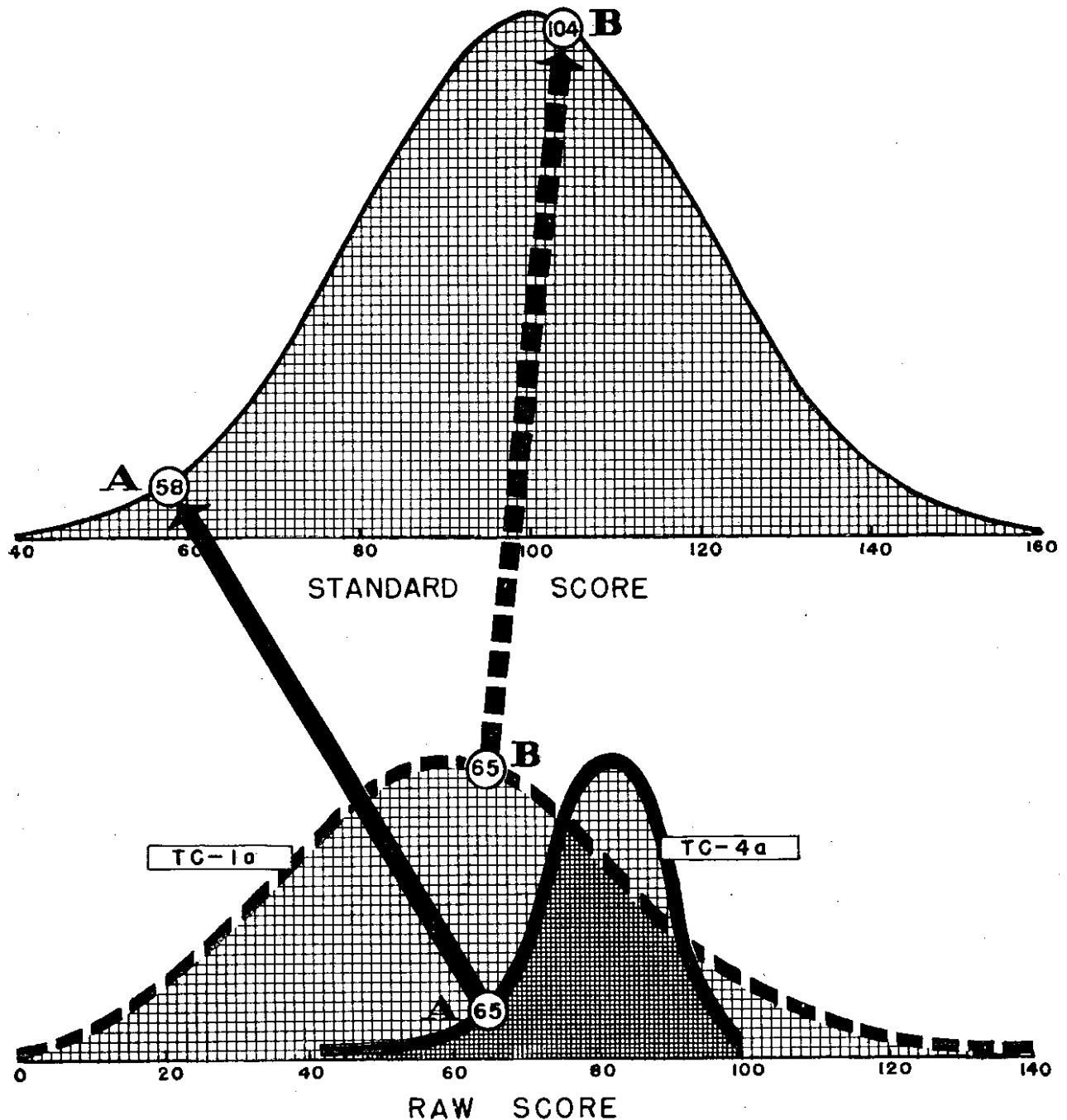


Figure 21. Conversion of raw scores to standard scores. Raw scores are not comparable. For example, a raw score of 65 on the Trade Information Test (TC-1a) is above the average; the same raw score on the Cryptography Test (TC-4a) is very low. This is shown clearly when both raw scores are converted to Army standard scores.

smallest in Grades I and V. It must be borne in mind that the divisions between Army grades are arbitrary; they may separate men of very similar abilities. For example, the difference between a man whose score places him barely within the Grade

III range will differ more from a high Grade III man than from a man near the upper limit of Grade IV. Moreover, a great many men belong in several different grades since they have more skill and aptitude for some assignments than for others.

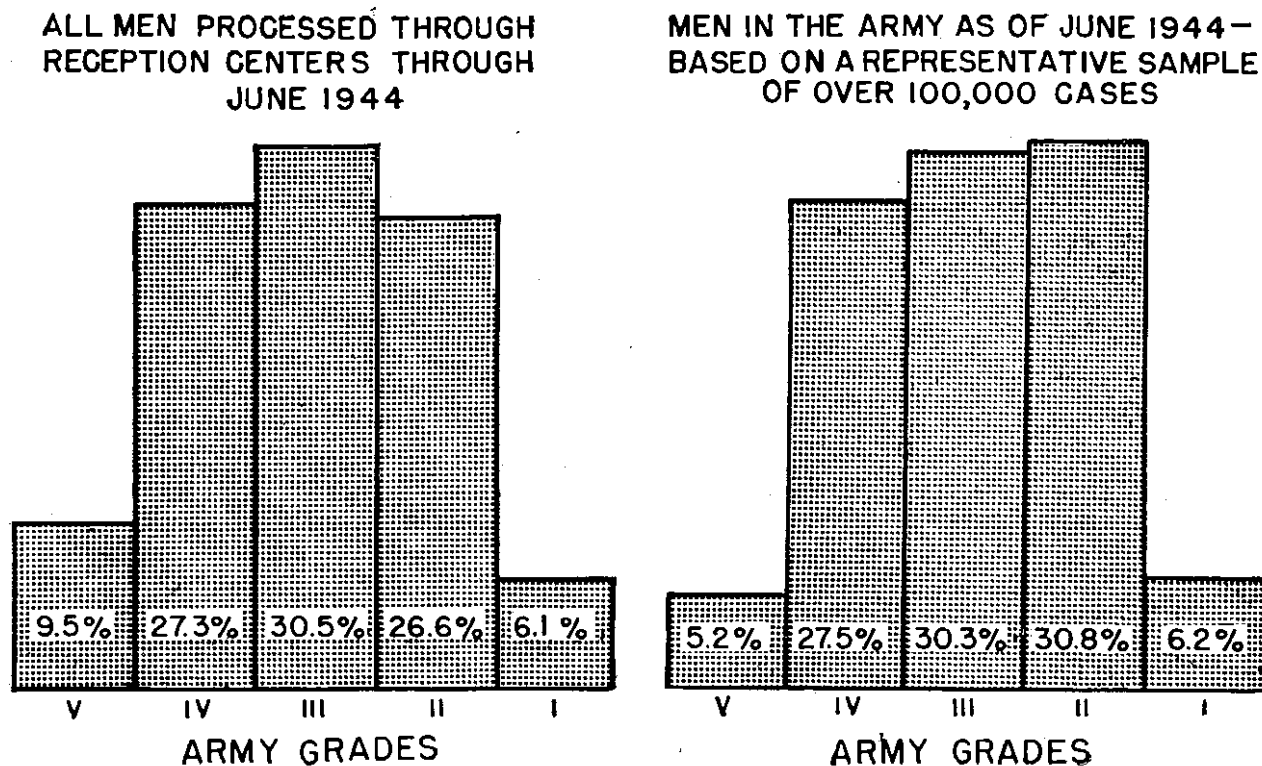


Figure 22. Army grade distributions of AGCT scores. The first figure shows the distribution of men taken into the Army through June 1944. The second figure shows the distribution of scores of men in the Army as of that date. The differences in these two figures are in part accounted for by discharge of men in the lower grades.

It is, therefore, a serious error to think of men as "Grade V," "Grade III," etc., except when they are being considered for a particular assignment in terms of the tests applicable to that assignment. It is equally erroneous to roughly and subjectively "average" a man's scores on different tests and thus classify him in a hypothetical, composite grade which has been neither developed nor authorized by the Army. A great many assignments require specific skills and aptitudes in the degree designated by directive, and performance on other irrelevant tests cannot be substituted.

b. Another type of score sometimes employed to express test results is the *percentile score*. A percentile may be defined as a figure expressing the percentage of scores in the population being considered which is exceeded by the particular score for which interpretation is sought. Thus, a score which equals or exceeds 90 percent of all scores made by men tested for radio code-learning aptitude is a score at the 90th percentile. The relation between percentile scores and Army standard scores and grades is illustrated in figure 20.

#### 64. Army Standard Scores Are NOT "IQ's"

Army standard scores bear no relationship to such concepts as the "IQ" (Intelligence Quotient), or "MA" (Mental Age), and Army test results will not be interpreted in terms of these concepts. (See par. 19, TM 12-425.)

### Section II. SIGNIFICANCE OF TEST SCORES

#### 65. General

a. To grasp the full significance of a test score, it is necessary to understand something more than the fundamental basis of the Army standard score scale. That a soldier may rank in the top 10 percent of all Army men in performance on a particular test is factual information. That he has brown eyes and wears a size  $7\frac{3}{8}$  hat is also factual information. But the Army is not concerned with the collection of odd scraps of biographical data. Eye color and head size, to be sure, may be valuable clues to identification. But a test score, whether high or low, remains an isolated and useless fact

until its significance is fully realized. And the first step in realizing this significance is to know what the test measures.

b. It is impossible to assign all Army tests titles which give a clear understanding of what is measured by them. The Clerical Aptitude Test measures aptitude for clerical work; and the Cryptography Test evaluates either aptitude or skill in cryptography. But one may well ask: What about the Army General Classification Test? For what does a high score on the Qualification Test qualify? Titles that seem the most specific are apt to be the most misleading. Clerical assignments, for example, are of many kinds and varieties; the Clerical Aptitude Test does not have equal significance for them all, and for some has no predictive value whatever.

## 66. Test Scores Predict Performance

a. It has been repeatedly emphasized that Army tests are designed to meet specific selection problems. It has been shown that each test measures certain skills or abilities because these skills or abilities are essential for successful completion of training courses or adequate performance of duty assignments. Army tests are used, therefore, because they *predict performance* in advance of selection or assignment. To comprehend the full meaning or significance of a test score, then, it is necessary to

understand precisely *what* performance the score predicts, and *how well* it predicts that performance. Both of these questions can be answered only after field trials and statistical analysis of the results.

b. A review of the manner in which a test is developed (ch. 3) will give a partial answer to the question of what kind of performance the test score predicts. This review will show that test items are selected because they are related to particular aspects of performance on certain job assignments—related, that is, in the sense that the items are answered correctly by the same persons who rate high in those aspects of performance on the job. It will be recalled that these phases of job success that are employed as standards of comparison in the selection of test items are called *criteria* of success. Since the test was built in this fashion, it can be assumed that each individual's score will predict his standing on the particular criteria of job performance employed. For example, if a test is constructed by selecting items which prove to be related to code-receiving speed, then it can be shown that scores on the test will predict speed in receiving code after a certain period of training. It should be noted, however, that the test cannot be assumed (in the absence of proof) to measure any other criterion of radio operation—though this often proves to be true. This is one reason why the review of the paragraphs on test construction gives only a partial answer to the question of just what performance the test score will predict. Another reason is that some tests, like the Army General Classification Test, are not designed to measure a single skill. Such tests can be used to evaluate chances for success in a large number of jobs or training courses which involve “general learning ability.” They may do a better job of predicting performance in some of these cases than in others. So the best and most logical way to determine what a test measures is to make a comparison between the performance of men on the test and performance of the same men on actual assignments or in training courses. The statistical statement of this comparison, then, will indicate how well the test performance predicts job performance. If those men who score high on the test are also successful on the job, the test is said to be a *valid test for purposes of predicting performance on that particular job*. If the relationship is a consistent and close one, the test is said to have *high validity* for that purpose. *It is important to recognize that the same test may have high validity for some purposes, low validity for others,*

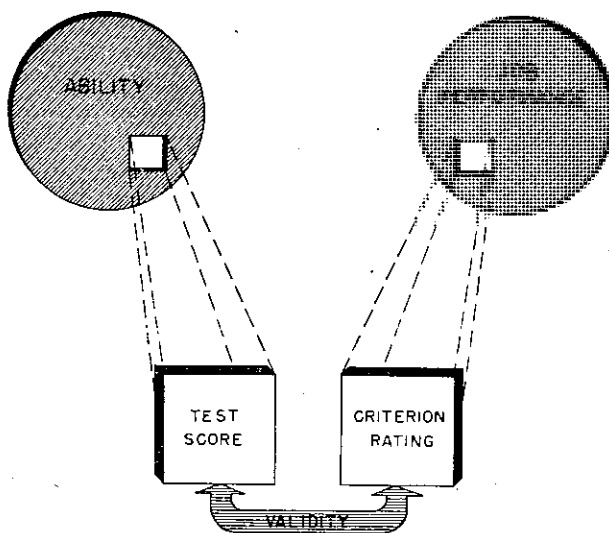


Figure 23. The test samples the abilities of the individual and the criterion rating estimates performance on the job. The correlation between the test score and the criterion rating thus indicates the accuracy with which the test score will measure a soldier's ability to perform a job.

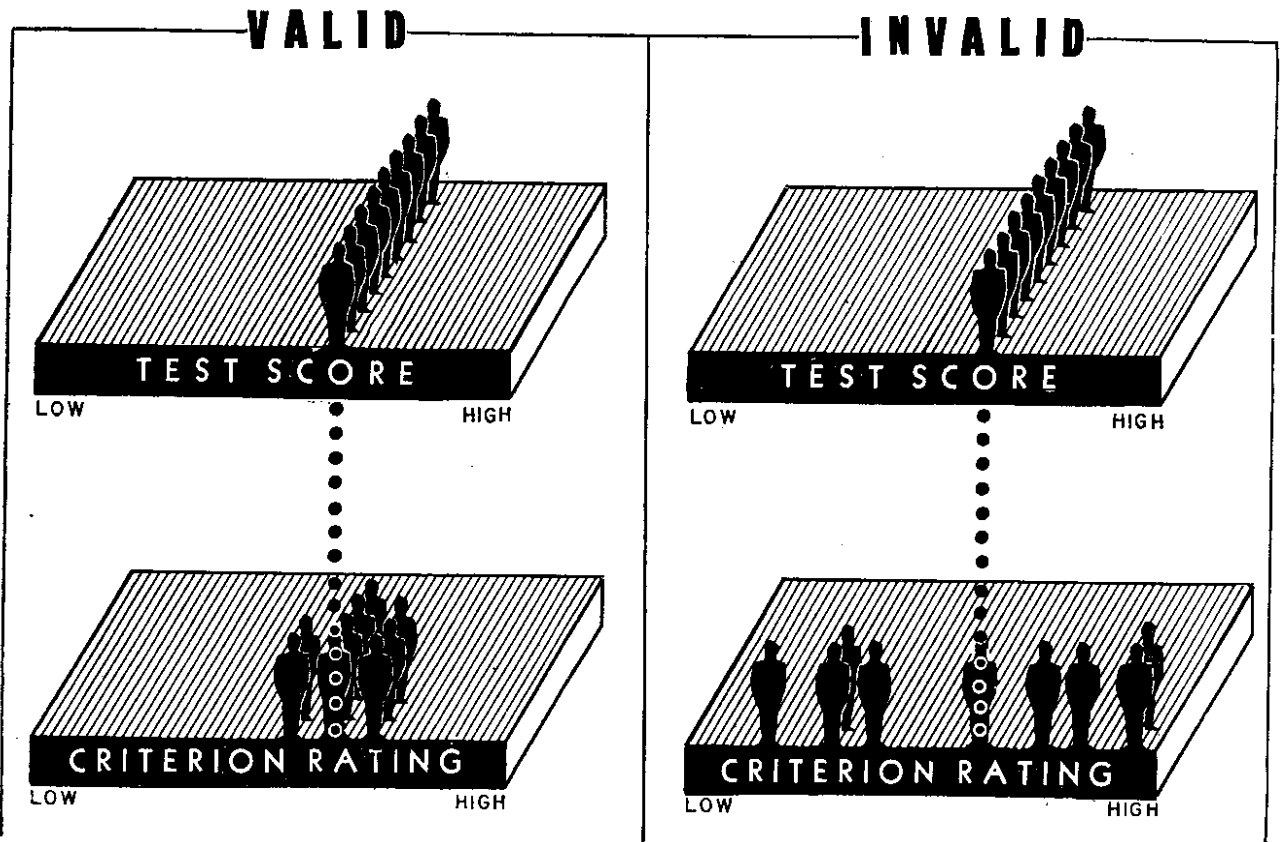


Figure 24. Test scores are employed to predict criterion ratings. The more nearly the men who get the same test score tend to achieve the same criterion rating, the more valid the test, and the more accurate the prediction.

and under other circumstances, no validity at all. Before employing any test result for assignment to any job or training course, the classification officer should satisfy himself that the test is valid for that purpose. Test validities are specific; they are evidenced, in each case, by demonstrated relationships between test scores and criteria of job success.

## 67. Validity Coefficient

The degree to which performance on the test compares with performance on the criterion can be portrayed by plotting both on a graph.

a. The process is somewhat analogous to plotting a distribution of scores; but since there are two scores for each person (test score and criterion score), it might be termed a double or square distribution. The way in which this graph is constructed is shown in figure 25. The mathematical statement of this relationship between test score and criterion score is called the *validity coefficient*.

b. The validity coefficient is extremely valuable. It tells, first of all, whether the test has enough validity to be useful. In general, it can be said that

any test which saves the Army time, materials, manpower, and money is sufficiently valid so long as it cannot be replaced by a better test. Often a test of comparatively low validity is preferable to the alternative of purely chance or random selection and assignment. Secondly, the size of the validity coefficient makes it possible to predict the approximate criterion score of men receiving any given test score. Thus, from a knowledge of a man's score on a valid test, it is possible, within the limits of probability, to predict how successful he will be on the job, or to compute his chances of becoming average or better on the job or training course.

## 68. Expectancy Charts

The research work and statistical computations involved in determining the validities of tests for various duty assignments is intricate and requires long training. It is not expected that this can be done by classification personnel in the field. Much of this work is executed by personnel procedures officers and expert technicians, and the more important results are made available in readily in-

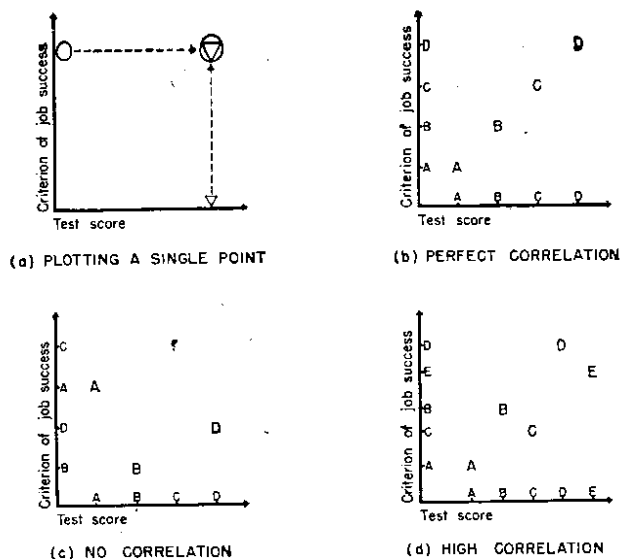


Figure 25. Test scores plotted against criterion scores to show degree of correlation. In diagram (a), one point is plotted. The triangle on the line labeled 'test score' represents the test performance of an individual. The circle shows the criterion score of the same man. The dotted lines are drawn perpendicularly to the two base lines at these two score points. These dotted lines intersect in a single point which represents the individual's score on both test and criterion. The remaining diagrams show graphs with 4 or 5 points plotted. In (b), where the four individuals each have the same rank on both the test and criterion scores, it can be seen that the plotted points all fall on a diagonal straight line. This indicates a 'perfect' correlation. In (c), there is no relation between the two sets of scores, and the plotted points form more or less of a circle. This indicates a very low or 'zero' correlation. Usually the degree of relationship will be somewhere between these extremes and the plotted points will form an ellipse as in (d). The narrower this ellipse, the more it approaches a single line, the higher the correlation. The wider the ellipse, the more it approaches a circle, the lower the correlation.

terpreted and usable form. For the most part, these data are presented in the form of *expectancy charts*. A number of these charts are provided in chapters 8 and 9 of this manual, and their use is discussed more fully there. In effect, the expectancy chart reveals in graphic terms the expected job performance of men receiving various test scores. More specifically, they show the probabilities (chances in 100) of better than average success on particular job or training assignments for men receiving various scores on particular tests. Their importance in connection with the use of tests in classification and assignment should be obvious.

## 69. Critical Scores

a. For purposes of assignment, field work and statistical analysis may result in the stipulation of

certain *critical scores* which must be achieved or surpassed by every man selected. As stated previously (par. 37), critical scores are set at a point dictated by practical necessity. If the supply of men is large in relation to the demand, the critical score may be set so high that nearly all the men selected will succeed. In other words, Army tests may predict success very accurately if the critical score is set high enough. The manner in which critical scores are adjusted to meet Army needs is illustrated in figure 26.

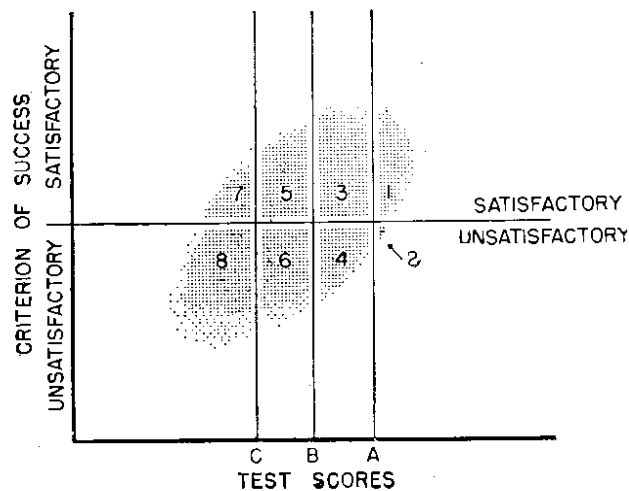


Figure 26. Proportions of satisfactory and unsatisfactory men who would be selected by various critical test scores when the validity is fairly high. If the critical score is set at A, only the men represented by the areas 1 and 2 will be accepted, and nearly all of these will be successful (1). If more men are needed, the critical score may be lowered to point B. Then an added group of men (areas 3 and 4) will be accepted. Most of these will be successful (3), but an appreciable number will fail (4). If the critical score is lowered still further (to C), the additional men admitted (areas 5 and 6) will be about equally divided between satisfactory and unsatisfactory. Lowering the critical score inevitably means that a larger number of failures will be selected, but the Army must sometimes countenance this waste in order to obtain the requisite number of qualified men. It should be noticed that even when the critical score is lowered to C, the total number of satisfactory men selected (areas 1, 3, and 5) is considerably in excess of the total number selected who are unsatisfactory (areas 2, 4, and 6). The critical score is often called the 'cut-off point,' since men who score below it are not considered for the assignment in question.

b. The situation arising from selection on the basis of critical scores is often complicated by the fluidity of standards of success. Since men are usually considered satisfactory if they are average or better and unsatisfactory if below average, it follows that standards will be adjusted to the average level of the group. When critical scores are lowered, thus lowering the average level of the



group, standards of performance on the assignment tend to decline and a larger-than-expected number of men will be considered satisfactory. Conversely, when critical scores are raised, standards of performance tend to rise and fewer men will succeed than the test scores would indicate. The *exact* figures in expectancy charts, therefore, should be expected to apply only when course or job standards remain fixed. (See par. 119)

### Section III. RELIABILITY

#### 70. Definition

An ideal measuring instrument would yield equally accurate results every time it was used. Such an instrument would be perfectly *reliable*. This ideal is often approached, but never, even in physical measurement, completely attained. There are some imperfections in every instrument. Moreover, there are exceedingly small but discernible changes in the materials to which the instrument is applied—in the case of physical measurement, changes resulting from variations in temperature, moisture, pressure, etc. With psychological measuring instruments, the situation is somewhat aggravated. The test itself often falls farther short of perfection than is usual with physical instruments. But the variations in the men measured, due to such influences as fatigue, previous testing with the same or a similar test, lack of interest or effort, and variations in the manner of administering the test from time to time, or examiner to examiner, contribute most to the unreliability of test results. These latter factors, which originate outside of the test itself, can be largely controlled by a strict adherence to the regulations and directions concerning the administration of all Army tests. (See ch. 4.)

#### 71. Estimating Reliability of a Test

a. **GENERAL.** Allowance must be made, however, for the small proportion of those factors extraneous to the test which cannot be controlled, and for the imperfections of the test itself which make for less than perfect reliability. It is thus necessary to determine just how reliable the test is in order to know how much unreliability to allow for. Army tests are subject to rigorous and exhaustive statistical checks before being released to

determine the accuracy of the measures which may be obtained with them.

b. **METHOD.** Comparing the scores achieved by men who take the same test two or more times under identical circumstances, or who take exactly equivalent forms of the test at one time, would show whether the test is reliable. For, if the same man gets the same score each time or on each equivalent form, the test must certainly be reliable. Whatever trait it measures, it measures consistently; but if many individuals get widely different scores each time, it must be unreliable, and no confidence can be placed in any single score. It is impossible, however, to give the test more than once under identical circumstances. Besides the obvious effects of familiarity, there are bound to be other changes in the examinees from time to time. And it is usually too difficult (if not impossible) to construct an exactly equivalent form of the test. In practice, therefore, the Army makes use of a statistical technique\* which estimates the result that would obtain from the administration of exactly equivalent forms of the test at a single session. The technique is accurate for most types of tests and has the important advantage that it saves both time and effort. The mathematical statement of the result is called the *reliability coefficient*.

#### 72. Use of Reliability Coefficient

The reliability coefficient gives important information about the test and the interpretation of test scores.

a. **IMPROVING TEST RELIABILITY.** The first use of the reliability coefficient is to determine whether the test is reliable enough for classification purposes. If the reliability proves to be too low, it can usually be increased by the addition of more items of the same kind. (See par. 25.) But since long tests are time-consuming, reliability beyond practical usefulness is not sought. All Army tests, before being released, are made sufficiently reliable for use in classification and assignment, provided they are used as directed in the manuals accompanying them and for purposes for which they were designed.

b. **INTERPRETING TEST SCORES.** (1) *General.* Because Army tests are reliable instruments, it is improbable that an individual's score will vary enough to present a practical problem in classi-

\*Kuder, G. F. & Richardson, M. W. The Theory of the Estimation of Test Reliability. *Psychometrika*, 1937, 2, 151-161.

fication work. At the same time, it should be recognized that very small differences between scores (5 or 6 standard score points, for example) should seldom be interpreted as reflecting real differences between the individuals receiving them. This statement does not imply that, where critical scores are stipulated, persons scoring only 5 or 6 points below these standards should be considered as passing. While it is true that a number of such people, if retested, might very well reach or exceed the critical score, it is equally true that a like number will score lower than before. It will be recalled that critical scores are computed on the basis of quotas and manpower supply as well as expectancies of success. Any laxity in applying these standards, such as that involved in the practice of making exceptions in individual cases that are "close," can only upset quota calculations without at all increasing the validity of selection.

(2) *Retesting.* Because Army tests are reliable, it cannot be expected that retesting, by and large, will result in marked score increases. Familiarity with the test and the test situation may contribute a few points of increase, but experience proves that this increase, on the average, will be small. Moreover, it is important to recognize that there is no virtue in getting high scores for their own sake. The only purpose in using tests at all is to enable predictions of job or training success, and there are seldom grounds for supposing the higher retest score to be a better predictor than the original—on the contrary, there usually is reason to suspect its validity. Therefore, except where there is marked discrepancy between a man's test score and his abilities as inferred from other information (for example previous education and occupation) the practice of retesting should be discouraged. Re-

testing is governed by the regulations contained in paragraph 18, TM 12-425.

## Section IV. SUMMARY

### 73. General

The following information is provided to aid classification officers and other personnel in the correct interpretation and use of test data.

a. INFORMATION PROVIDED IN TEST MANUALS ISSUED WITH EACH TEST. (1) *Conversion tables* which enable field personnel to translate raw score into standard scores. The latter show the comparison of each tested individual with all other Army men as to the particular skill or aptitude which the test measures.

(2) *Validity coefficients* given in terms of the available criteria. These indicate the degree of probability that a man who achieves any given Army standard score will achieve a certain predictable criterion score. That is, in more general terms, they show how accurately standard scores on a test predict success on the designated assignments.

(3) *Reliability coefficients* which indicate the accuracy with which the test measures every tested man.

b. INFORMATION PROVIDED IN DIRECTIONS. *Critical scores* which must be achieved or exceeded by men assigned to specified training courses and duties.

c. INFORMATION PROVIDED IN THIS MANUAL (ch. 8) *Expectancy charts*, which show the chances in 100 that men who achieve designated test scores (ranging from low to high) will achieve average or better performance on specified assignments. Expectancy charts are provided with statistical data requisite to their proper interpretation.

## CHAPTER 6

# RATING SCALES, QUESTIONNAIRES, AND INTERVIEWS

### Section I. ORIENTATION

#### 74. General

The tools of the Army classification system are techniques or instruments designed to aid in the evaluation of the characteristics of men in the Army. They are techniques employed to discover and to measure as accurately as possible what each man can do or is capable of learning to do. Whenever possible, scientific measuring instruments are used. (See ch. 2.) These instruments are the numerous tests that are described elsewhere in this manual and that have proven of unquestionable usefulness in the evaluation of skills, capacities, and aptitudes.

a. There are, however, characteristics of men that enter into the determination of success in training or in assignment—personality traits and social adjustments which do not lend themselves as readily to measurement by tests. For example, no instrument has as yet been devised to test the traits of “leadership” in a satisfactory manner. Yet most men would agree that individuals differ with respect to the degree to which these traits are developed; and further, that these differences in “leadership,” properly rated, are important factors in classification and assignment.

b. In those areas in which tests are available for evaluating skills and capacities, additional evidence is often useful to round out the data on individuals and thus make sound classification and assignment doubly sure. Information about previous occupational experience, education, interests, and hobbies is of this nature. And such information is obtainable in useful form by a proper employment of the interview, the questionnaire, and the check list.

#### 75. Purpose and Scope of This Chapter

This chapter is included to orient the testing officer in the other procedures used in the Army classification system, so that he will understand their relationship to testing as conducted in the Army, and the important parts played by each of the tools used in personnel evaluation. The techniques and methods described here are, of necessity, less formalized than tests. A few special questionnaire forms

and check lists have been developed and authorized for classification purposes. (See ch. 8.) The interview is employed not only in initial classification in the reception center (ch. 7), but in all echelons of the Army. It is the purpose of this chapter to describe and evaluate these tools, and to provide suggestions that will aid in their development, use, and interpretation. The judicious use of these devices, together with the information derived from tests, will achieve a well-rounded program of classification and assignment.

#### 76. Value of Rating Scales

A good rating scale enables one person, on the basis of adequate observation, to judge another and to assign to him quantitative evaluations of characteristics not at present amenable to scientific measurement through tests. Examples of such characteristics are physical appearance, bearing, leadership, manner of speech, etc. The rating scale also provides a convenient means of obtaining estimates of proficiency or manner of performance in duty assignment when the finer discrimination provided by a test is either not feasible or not necessary for practical purposes.

#### 77. Value of Questionnaires and Check Lists

Carefully constructed questionnaires and check lists are useful instruments for obtaining information about the background of the individual, his educational and occupational training and experience. They may also be used to make inventories of personal characteristics such as interests, attitudes, and social reactions. Questionnaires of this sort are often elaborate devices that can be “scored” like a test.

#### 78. Value of the Interview

The interview is the source of very important information: as such it is an oral questionnaire. It may be used to coordinate and evaluate data from several sources according to its impact on the individual. It may be used as a medium for giving information, advice, guidance, or therapy. It may serve all of these purposes at the same time, and it may further, because of its essentially intimate

**RATING SCALE OF COMPETENCE IN PERFORMING A WORK PROJECT**

NAME \_\_\_\_\_  
Last First Middle  
ASN \_\_\_\_\_ CLASS \_\_\_\_\_

Instructions to rater: Consider carefully each of the five descriptive paragraphs below; then, on the basis of your observation of the trainee whose name is entered above, decide which of these paragraphs best describes his work on the project and place a check-mark (✓) in front of that paragraph.

- ☐ (1) Could not complete job even with major assistance from instructor. Did not know the relative parts of his job either by definition or use. Had no understanding of why the job was to be done.
- ☐ (2) Was able with difficulty to complete parts of the job himself. Had an idea what to do but lacked sufficient information or dexterity to complete all parts of the job. Understood very little of why he did the job.
- ☐ (3) Had a general idea of what was to be done. Finished the job but with minor errors of omission or commission. Made false starts, changes, and repetitions. Was not sure of himself or his product.
- ☐ (4) Completed the required job with little hesitancy. Learned what to do and understood generally the underlying principles.
- ☐ (5) Completed the job quickly and efficiently. Learned what to do, why to do it, and the relationship of this job to others being studied in the unit.

RATER \_\_\_\_\_

RANK \_\_\_\_\_ ORGANIZATION \_\_\_\_\_

Figure 27.

character, add the touch of personal attention that can do so much for morale.

## 79. Comparison with Tests

Interviews, questionnaires, check lists, and rating scales are valuable personnel instruments. As devices for obtaining accurate and impartial information about abilities and aptitudes, they are inferior to tests. As techniques for guiding and objectifying informal data-gathering, they are far superior to unaided human judgment. To be of any real use, the questionnaire and rating scale must be worded and constructed with care. And if it is to serve any of its ends, the interview must make use of men who are friendly and patient and well trained in the arts of questioning and counseling. At their worst, these techniques can introduce bias and error and decrease the validity of evaluation and classification. At their best, they can be extremely useful. But they should never be considered as substitutions for the scientific measurement by tests wherever tests have been designed for specific purposes.

## Section II. RATING SCALES

### 80. Definition

The rating scale may be cast in several forms. The proper form should be chosen according to the nature of the problem to be solved, the trait or traits to be evaluated, and the persons who are to make the ratings. Despite minor variations and elaborations, the rating scale is essentially a device for eliciting evaluative judgments of traits or qualities of individuals in such a manner that these judgments can be handled quantitatively. Evaluative judgments of traits or characteristics are constantly being made in all kinds of situations. The soldier makes informal judgments about the abilities of his fellows or the competence of his officers. The squad leader, sergeant, platoon leader, and company commander are constantly evaluating the military proficiency of their men. Commanders of higher echelons render periodic efficiency ratings on their officers. And, in schools and training centers, instructors pass judgment on the progress and competence of their trainees. Most of these judgments are couched in qualitative terms, like "good" or "average," "industrious" or "lazy." Such ratings, like the free-answer or essay type test response, are highly subjective, open to individualistic interpretation, and difficult to handle

in meaningful fashion. The rating scale provides an objective framework for the evaluation. It offers the rater a nicely graded series of guideposts within which he may indicate his judgment by making a check mark or assigning a predetermined numerical "score." Figure 27 is an example of a simple type of rating scale for evaluating the general competence of trainees in their performance of a specific task.

### 81. Types of Rating Scales

The rating may take several forms.

a. **RANKING.** An obvious method of rating a group of individuals according to their possession of a particular trait or characteristic is that of ranking. This consists of assigning the number 1 to the person who stands highest in the group with respect to that characteristic, the number 2 to the person who stands second, and so on down through the whole list. Although this method is easily understood, it has serious disadvantages. It is exceedingly cumbersome when the number of persons to be rated is large. Moreover, it is impossible to combine ratings on groups of various sizes since the numerical value assigned to each person—his position in the group—obviously depends on the size of the group. For example, the poorest person in a group of 5 receives the same "score" as a man in the top fourth of a group of 20. And again, the standard of reference of the ratings with this method is always the average of the group being rated rather than the average of all trainees, of all men in that assignment, or of all men, in the Army.

b. **NUMERICAL RATING SCALE.** Some of the difficulties encountered in the method of ranking are remedied by means of the numerical rating scale or the scale of values. With this method a number of categories are set up in descriptive terms and assigned predetermined numerical values. Each of the persons rated is then placed in that one of the categories which best describes him, and given the numerical value or "score" that goes with it. Figure 27 is an example of this type of rating scale. Whether or not groups rated separately can legitimately and meaningfully be combined with this method depends upon whether or not all raters interpret the descriptive paragraphs in the same manner. In general, the reliability of the rating scale is impaired because the standards of different raters vary.

STUDENT OFFICER GRAPHIC RATING SCALE					
Name _____ (Last Name) (First Name) (Middle Name)					
Military Bearing and Neatness	Slovenly Not Neat	Poor Posture Physical Defects	Clean, Orderly Average	Very Neat Well-Groomed	Prepossessing Attracts Attention
Physical Adaptability	Abundant Energy Athletic	Well-Coordinated Vigorous	Average Does Not Attract Attention	Exerts Himself with Effort, Listless Awkward and Clumsy	Lazy, Stands on Sidelines
Group Adaptability	Repellent Is Avoided by Others	Colorless Does Not Attract	Accepted By Group	Congenial Is Sought After	
Attitude	Dynamic Self-Starter	Determined Resourceful Enthusiastic	Interested Shows Desire to Succeed	Interested But Requires Supervision	Not Interested Resistant
Ability to Express Himself	Inarticulate Illiterate	Has Difficulty In Expressing Himself, Uses Bad Grammar & Language	Gives Directions Well	Clear Direct	Convincing Forceful Persuasive
Ability to Handle Men	Inspires Others Makes Things Go	Can Organize And Direct Others	Takes The Lead Sometimes	Depends On Others	Never wins Group Support, Creates Antagonism
Stability	Easily Moved To Anger, Fits of Depression, Blows Up	Lacks Poise Gripes Gloomy	Occasionally Gripes	Well-Balanced Stands Fast	Meets Emergencies Calmly
Remarks _____					
Period: First To _____ Week, Inclusive			SIGNED: _____ (Signature) (Grade) (Branch)		
Date: _____			Commanding _____ Company		

Figure 28.

c. GRAPHIC RATING SCALE. If the descriptive categories are represented as distances along a line, the scale becomes a graphic rating scale. Figure 28 illustrates the technique applied to some seven characteristics that are considered important for officer candidates. The amount of each characteristic is represented as a distance along the line with descriptive phrases acting as milestones. The rater places a check mark at any point along the one which indicates his judgment.

## 82. Construction of a Rating Scale

The construction of a rating scale, like the construction of a test (ch. 3), should start with a clear-cut formulation of the purpose to be served by the scale. This statement of purpose should encompass the nature of the selection problem which demands the evaluation of traits, the specific trait or traits to be evaluated, the population to be rated, and the persons who are to utilize the scale as raters. The

exposition that follows has particular reference to the numerical scale and the graphic rating scale.

a. **DEFINITION OF CHARACTERISTIC TO BE RATED.** To insure that all raters gain as nearly as possible the same understanding of just what characteristic of the individuals they are to judge, it is necessary to have a clear-cut and complete definition of this characteristic. If, for example, officers are asked to rate men with respect to "cooperation," the task is difficult since each officer will have to decide for himself just what constitutes this trait. The description of the trait should be in terms of concrete behavior that can be observed rather than abstract qualities. The definition should thus be based on a "job analysis" and on observations of individuals displaying the quality to be evaluated. The following is an adequate definition of cooperation:

*Cooperation.* Consider how well the soldier works with his supervising noncommissioned officers and other soldiers. Take into account whether he readily assumes assigned duties and responsibilities which may be inconvenient, whether he freely offers to help, how well he works at jobs requiring teamwork, etc.

b. **NUMBER OF DEGREES IN SCALE.** The next step is to decide how many degrees or steps the scale is to have. This will depend upon how well the trait is defined, how precisely and objectively it can be observed, and how many degrees of it can be discriminated. In some cases it will not be worthwhile to attempt more than a three-step discrimination such as "above average," "average" and "below average." Usually finer discrimination into 5 or even 7 degrees is possible.

c. **DEFINING POSITIONS ON SCALE.** In both the numerical or the graphic type of scale, the various scale categories must be clearly defined. These are the guideposts of the scale, and like all guideposts, should be explicit and unequivocal. They will achieve this necessary clarity if they are described in terms which refer to objective behavior, to activities of individuals displaying the varying degrees of the trait. With respect to the characteristic of "cooperation" cited above, for example, the following five categories might be used to define positions in the scale:

(1) Frequently balks at doing inconvenient tasks; never puts himself out to help others; frequently in disagreement with other soldiers.

(2) Avoids doing inconvenient tasks; does not get along well with other soldiers.

(3) Carries out most duties and responsibilities assigned to him, although he occasionally tries to avoid inconvenient assignments; works with others fairly smoothly.

(4) Readily assumes all duties and responsibilities which may be assigned to him; gets along well with others.

(5) Goes out of his way to be helpful; volunteers for duties and responsibilities; works very well with others.

d. **"HALO" EFFECT.** When a rater evaluates an individual with respect to a number of different traits, he is likely to rate in the same way on all of them regardless of their independence. Having formed a general impression of "excellence," for example, he will tend to credit the individual with above-average standing on all of the characteristics included in the scale. This tendency to base ratings on a general overall impression, rather than on an independent consideration of each characteristic in turn, is called the "halo" effect. It must be avoided if the separate ratings are to have maximal usefulness. There is no reason why all of the characteristics in the scale illustrated in figure 28 should necessarily be related. It is quite possible for a person to be able to handle men well, to express himself with facility, and yet be untidy in his dress and bearing. Yet if the officer using this scale has formed a general impression of the individual as an undesirable candidate or unlikely officer material, he will tend to place all of his check marks toward the lower end of the scale if desirable and undesirable characteristics are listed in columns. It will help to avoid this tendency to vary the alignment of the scale in such a way that the high or "good" extremes fall as often at the left as at the right ends of the lines. This arrangement may help to force the rater into considering each scale separately; but the best way to avoid the effect of the "halo" is through the training of raters.

### 83. Preliminary Try-out of Scale.

After the scale has been constructed in preliminary form, it should be subjected to a thorough try-out under real conditions. It can then be revised and adjusted in the light of this experience. The rating scale, like the test, should be checked for reliability and validity. Its reliability can be tested by determining the extent to which soldiers in a group are rated in the same way by independent raters, or by determining the consistency of the ratings made by a single judge on different occasions. The

validity of the scale can be computed in the normal manner by determining the relationship between the ratings of a group and some independent criterion of success in training or on assignment. It should be evident that any rating scale, no matter how carefully constructed or how reliable it may be, is of little value if it does not provide evaluations that are related to success in training or on the job. The preference for measurement by tests, where these are available, is largely based on the fact that the reliability and validity of rating scales are generally somewhat lower than those of tests.

#### 84. Administering and Scoring Rating Scale

a. Abundant experience has demonstrated that ratings which are the result of the combined judgments of several individual raters are more reliable and valid than those obtained from single raters. It is, of course, obvious that the several ratings should be completely independent and not the result of previous discussion and collaboration among the judges. At least three independent ratings of each individual on each trait should be obtained if feasible. These can then be averaged to provide the best estimate of the amount of each trait the individual possesses.

b. (1) Those persons who are to act as raters must be thoroughly oriented and trained if their evaluations are to be of maximal usefulness. The orientation should cover such factors as the purpose of the ratings to be obtained, the kinds of behavior to be observed, the avoidance of the "halo" effect, and the necessity for uniformity in the interpretation of the various traits and the various positions on the scales.

(2) The desired uniformity of standards can best be achieved if the raters bear in mind the approximate nature of the distribution of ratings that is desired. On a five-step scale, approximately 40 percent of a large unselected group of individuals would be expected to be rated in the "average" or central position, with 23 percent just above and just below this position and 7 percent at each extreme. Untrained raters should be given practice in observing behavior, evaluating traits, and apportioning ratings.

c. "SCORING" RATING SCALE. It is convenient to refer to the numerical value assigned to a rating as a "score." The process of "scoring" a rating scale, therefore, is that of determining what that value should be. There are many possibilities. The simplest is to use the serial number of the

category checked. If the third, or middle, category of a five-step scale is checked, for example, the "score" on that particular category would be 3. With the graphic rating scale, the rater may place a check-mark at any point along the line, and consequently the "score" can be the distance of this check-mark from the end of the line. By measuring this distance to the nearest inch or in any fraction thereof, any desired degree of discrimination may be "scored." As stated above, several independent ratings of the same trait may be averaged for greater reliability. Where several different traits are rated, as in the scale illustrated in figure 28, the separate "scores" may be combined to give an overall rating of "leadership" or proficiency. They may be combined by simple addition, or multiplied by weights and added, depending upon the use to which they are to be put and their correlation with the criterion of success. Or the separate "scores" on each of the different traits may be plotted as a "profile" to provide a graphic presentation. (See fig. 29.)

PROFILE OF STUDENT OFFICER RATINGS										
Instructions for filling out profile chart are given on back of card.										
Traits Rated	10	9	8	7	6	5	4	3	2	1
Military Bearing and Neatness										
Physical Adaptability										
Group Acceptability										
Attitude										
Ability to Express Himself										
Ability to Handle Men										
Stability										

NAME John T. Smith ASN 321924231 CLASS 32  
 AGE 25 EDUCATION Bryans AGCT 123 OOT 117

Figure 29.

### Section III. QUESTIONNAIRES AND CHECK LISTS

#### 85. Definition

When used as a classification instrument, a questionnaire is a series of uniform questions of proved effectiveness which are presented to an individual for the purpose of eliciting information about himself. It is not to be confused with a test though the latter is also in many cases composed of questions. A test elicits behavior which indicates the



degree to which the examinee possesses knowledge, skill, or aptitude required for a particular training course or job assignment. A questionnaire, on the other hand, is employed primarily to secure personal information concerning the individual's past experience or his present attitudes. A questionnaire requires the individual to report upon himself, rather than causing him to demonstrate his characteristics in a measurable fashion. Nevertheless, a properly constructed questionnaire may furnish information valuable in predicting future achievement. When this is the case, the questionnaire is a valuable addition to the tools available for classification and assignment.

## 86. Information Gained Through Questionnaires

The questionnaire is customarily used to collect two kinds of information about the soldier.

a. **OBJECTIVE BIOGRAPHICAL DATA.** This includes such factual or verifiable information as the soldier's age, schooling, marital status, occupational history, and experience. Some data of this type (education and training, for example) are quite obviously of value in selecting prospective candidates for further training or suitable duty assignments.

b. **SUBJECTIVE INFORMATION.** This includes opinions, interests, attitudes, likes and dislikes, prejudices, as well as habitual fears and worries, and characteristic modes of behavior. Information of this kind about an individual may throw light on his personality and aid in predicting how he will act in various critical situations, more or less independent of his knowledge or skill.

## 87. Kinds of Questionnaires.

The questionnaire may be either oral or written.

a. (1) The oral questionnaire is essentially a standardized interview. (See sec. IV.) The interviewer asks a series of specific questions and records the oral answers of the subject.

(2) The classification interview conducted in the reception center is an example. It follows that any questionnaire may be presented orally, if this is desirable.

b. (1) In the more formal written form, the questions are presented in booklet form and the answers are written by the individual either in spaces provided in the booklet itself, or on a separate answer sheet.

(2) The written questionnaire has certain important advantages over the oral type. In the first place, it insures more uniformity in the method of presentation. Secondly, it provides more privacy. In the very personal face-to-face interrogation, the soldier will sometimes be restrained by embarrassment from giving full and honest answers. This restraint will be particularly strong if the questioner is unskilled in the art of interviewing. Consequently, the written questionnaire has an advantage in that it does not require skilled personnel for its administration. Moreover, the written form enables the soldier to allot his own time and to work longer on those questions which require "thinking out," without feeling rushed (as with the oral type) by the patient waiting of the ever-present interviewer. And finally, it is an obvious advantage of the written type of questionnaire that it can be given to groups with all of the economy of time and effort that group administration permits.

c. The *check list* is a particular variation of the written questionnaire. It is composed of a series of statements or descriptions, rather than questions, and the soldier is instructed to check those which are applicable or pertinent to him. It is particularly useful in gauging the range of occupational experiences or the extent of emotional problems. Figure 30 shows a portion of a check list used to determine the extent of the individual's experience in clerical work in the Army.

## 88. Construction of Questionnaire

The development of a questionnaire for use as a classification tool should proceed along the same lines as the development of a test. (See ch. 3.) It should first be decided for what purposes the questionnaire is to be used. Then in the light of a job analysis, it should be determined what sorts of information are to be obtained—whether age, schooling, etc., or interests, attitudes, and opinions.

a. The questions themselves should be phrased in such a manner that they are clear and unambiguous. It is not sufficient, for example, to ask "How much education have you had?" For results that are definite and unequivocal, the question should be worded somewhat as follows: "What is the highest school grade that you completed?" Special care should be exercised to avoid certain types of questions:

(1) The "double-barrel" question. This is really a compound question which can seldom be answered in a simple yes-or-no fashion. The ques-

# ARMY TRADE SCREENING TESTS

## CLERICAL EXPERIENCE CHECK LIST

### TC-23ar

NAME \_\_\_\_\_ ORGANIZATION \_\_\_\_\_  
 (Last) (First) (Initial)  
 ARMY SERIAL NUMBER \_\_\_\_\_ GRADE \_\_\_\_\_ DATE \_\_\_\_\_

A. Below are lists of Army clerical operations. If you have performed the operation, and you can do a good job now "on your own," check (✓) the item. Leave blank those items with which you have had very little or no experience.

#### 1. SOURCES OF INFORMATION

- ( ) use Army Regulations
- ( ) use AR 1-5
- ( ) use TM 12-250 (Administration)
- ( ) use TM 12-255 (Administrative Procedures)
- ( ) use Virtue's "Company Administration"
- ( ) use FM 21-6 (Index to Training Publications)
- ( ) use War Department Circulars
- ( ) use TM 12-252 (Army Clerk)

#### 2. COMPANY RECORDS

- ( ) make entries in the Morning Report
- ( ) check entries in the Morning Report
- ( ) make entries in the daily Sick Report
- ( ) maintain Duty Rosters
- ( ) write Company Orders

#### 3. PERSONNEL ADMINISTRATION

- ( ) make entries in Service Records
- ( ) make entries in Soldier's Qualification Card
- ( ) make entries in Officer's Qualification Card
- ( ) prepare furlough forms
- ( ) prepare locator cards
- ( ) check rosters
- ( ) prepare extract copy of Morning Report
- ( ) prepare descriptive list of Absentee Wanted by U. S. Army
- ( ) prepare Emergency Addressee Cards
- ( ) prepare discharge certificates
- ( ) prepare Certificate of Service
- ( ) prepare Report of Separation
- ( ) prepare Extract of Service Record
- ( ) draft Special Orders
- ( ) prepare Notification of Discharge

#### 4. MILITARY DISCIPLINE

- ( ) use Manual for Courts-Martial
- ( ) prepare Charge Sheets
- ( ) prepare Court-Martial Orders

#### 5. CORRESPONDENCE AND FILING

- ( ) type military correspondence
- ( ) type indorsements
- ( ) prepare inclosures (letter)
- ( ) use military abbreviations
- ( ) prepare messageforms
- ( ) use company correspondence file
- ( ) use War Department decimal file
- ( ) use policy file
- ( ) use suspense file
- ( ) use 201 file
- ( ) distribute incoming mail
- ( ) prepare outgoing mail

#### 6. MACHINE OPERATION

- ( ) operate an adding machine
- ( ) operate a ditto machine
- ( ) operate a mimeograph machine
- ( ) cut stencils
- ( ) use stylus
- ( ) type by touch method
- ( ) type by "hunt and peck"

#### 7. SUPPLY

- ( ) prepare Statements of Charge
- ( ) prepare Reports of Survey
- ( ) prepare Property Issue Slips

#### 8. FINANCE

- ( ) prepare pay rolls
- ( ) prepare officers' vouchers
- ( ) prepare allotment forms
- ( ) prepare Application for Dependency Benefits
- ( ) prepare final statements
- ( ) prepare insurance applications
- ( ) prepare Soldier's Deposit Books

B. Give below any important information which has not been covered and which you feel is necessary for a complete picture of your clerical and administrative background.

WD AGO PRT-245

Figure 30.

tion "Are you afraid of lightning and the dark?" is a case in point. The individual who is filled with terror by a flash of lightning might have no fear of darkness and so he is unable to answer truthfully except in the negative. If both aspects of the question are important, they should be asked separately.

(2) The leading question. This is a question which suggests its own answer. A very obvious type of leading question is the "you are, aren't you" type. "You are often uneasy when speaking before a group, aren't you?" strongly suggests an answer in the affirmative.

b. In the written questionnaire, the items should be worded in such a way that they require a minimum of writing on the part of the person answering them. This serves not only to make the questionnaire easier to fill out, but also insures greater ease in "scoring" answers and reduces the element of subjectivity. When possible, the questions are so phrased that they can be answered with a check mark, as in the example.

(1) Does it bother you to have 

Yes	?	No
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

 someone watch you at work, even though you know you can do it well?

With questions of this sort, of course, separate answer sheets can be employed and the answers "scored" by machine. The item may, also, be phrased as a statement, instead of a question, and checked as true or false depending upon whether or not it applies to the person answering it:

	T	F
(2) It bothers me to have someone watch me at work, even though I know I can do it well.	::	::
(3) I do not try to correct people who express an ignorant belief.	::	::

### 89. Preliminary Trial of Questionnaire

After the questions have been collected or constructed, the questionnaire should be thoroughly tested in operation. It should be given to groups representative of those for whom it is intended, to discover whether the questions are worded clearly, unequivocally and in language that is not too difficult. Its value as a classification instrument should be checked, in much the same way as a test is validated. (See ch. 3.) The Clerical Experience Check List (fig. 30), for example, might be given to a group of successful company clerks and to a random selection of soldiers. Analysis of the check marks made by both of these groups could then re-

veal whether there were any items checked as frequently by the group of nonclerical soldiers as by the clerks. Such items would be of no discriminative value, and so could be deleted from the list. On the basis of such preliminary trials and analysis, the questionnaire might be revised, some questions omitted and others reworded.

### 90. Scoring Questionnaire

The results obtained with a questionnaire may be handled in several ways. Since the answers indicate information about the individual, it follows that this information should be treated in the fashion required to make it an effective aid in assignment. Such data as age, schooling, occupational experience, etc., may be useful in themselves. Sometimes each item can be assigned a weight and entered into a predictive formula—so many points for successful completion of a specialist course, and so on. With check lists, such as that illustrated in figure 30, a simple summation of the check marks entered may be considered as a "score." It may be more important to know *what* duties the soldier has performed rather than *how many*, in which case, a study of the list is required. Personality questionnaires, or inventories, are often "scored" for several different traits; a certain answer to a given question may count so many points on the "aggression" scale, for example, and have some other value on another scale. In general, it can be said that the way in which the questionnaire is to be used will determine the method of handling its results.

## Section IV. INTERVIEW

### 91. Definition

An interview is a conversation with a purpose. It is directed along certain channels as predetermined by the person conducting interview. It is essentially a controlled conversation but the amount of control may vary over a wide range. In its most restricted form it may be of the formal question and answer type where it is limited to a series of specific topics. At the other extreme, the interview may have all the appearances of a thoroughly free and informal chat touching upon many topics according to the fancy of the participants. In a skilled interview, however, this appearance of complete indirection will be only superficial. The interviewer while purposely creating an informal setting and permitting the conversation to run freely

will, by means of questions and comment, imperceptibly guide the discussion along the desired channels. Both extremes of the interview have their place in good personnel procedures. Where detailed specific items of information are desired, the formal question and answer type of interview serves quite adequately. The oral questionnaire, as described in section III of this chapter, represents the extreme in this form of interview. As a clinical and counseling device, the free and informal discussion serves more adequately. Most interviews are intermediate between these two extremes. Both question and answer techniques and guided informal conversation are employed. Such is the classification interview as conducted in the reception centers.

## 92. Purpose of Interview

The interview brings two persons face-to-face and permits information to pass between them in both directions. This permits personalities to be brought into play with the result that a number of purposes may be accomplished. The accomplishment of some of these purposes is possible only because of this interplay of personalities; and results which could not be obtained with the use of impersonal techniques are thus made possible. For the most part the purposes accomplished by the interview may be grouped under three headings.

a. **TO OBTAIN INFORMATION.** The interview is a valuable, and sometimes the only, means of collecting information from the subject. Its value in this respect varies with the kind of information desired. When it is used as a method of revealing facts about skills and abilities, every effort should be made to verify and expand the information obtained by means of more objective techniques. As a means of collecting such data as age, education, date of birth, etc., the interview can be as reliable a measure as are test techniques and objective questionnaires. One of its greatest values, and one where it has advantage over tests and objective questionnaires, is in the realm of personality. When undertaken by one competent in the technique of interviewing, it can cut through attitudinal and emotional knots to the motivational core of a maladjustment in a manner impossible with the more impersonal instruments.

b. **TO GIVE INFORMATION.** As a means of orienting the soldier, instructing him or offering advice, the interview is superior to the class or to the lecture since it allows for a presentation of ideas or

counsel tailored to individual dimensions. Although the interview is more time consuming than the other methods, the time is considered well spent in terms of the satisfaction which the person interviewed expresses because of the personal attention given him. The separation counseling interview as used in the Army with its aim of disseminating vocational information and guidance is an excellent example of this type of interview. It is well to remember, however, that any successful interview is informative to both parties and an interview at the reception center should not only result in the desired information for the interviewer but must satisfactorily answer the questions in the mind of the interviewee.

c. **TO BOOST MORALE.** Here again the interview is an excellent device. It is immeasurably easier in the face-to-face situation to influence attitudes and emotions than it is by a formal lecture. Any unit commander maintains the spirit and morale of his organization largely by the personal contacts he and his assistants maintain with the individuals composing the organization.

## 93. Reliability of Interview

Caution must be exercised in forming judgments from information obtained in interviews. Sources of error are inherent in the interviewer, in the person being interviewed, and in the relations between the two. For example, failure on the part of the interviewer to formulate the problem in such terms that the interview can contribute to its solution, may result in the collection of data not at all pertinent to the problem at hand. Reliability of data is always limited to the interviewee's knowledge, his memory, his attitudes, his ability to observe, by his understanding of what is wanted and by his verbal capacity for clear and accurate expression of what he knows.

## 94. Value of Interview

While the interview has its weaknesses and limitations, it has great value as a classification technique. Used in conjunction with more objective methods, the interview affords the opportunity to integrate fragmentary and disjointed information and to reconcile apparently contradictory data. While subjective in character, it serves to uncover sources of information which may be investigated further by the more objective methods. The ultimate use of all information whether obtained through interview or by objective techniques is to obtain an over-

all picture of the individual. In the hands of the qualified interviewer, the interview technique furnishes valuable information on which such judgments may be based.

### 95. Qualifications of Army Interviewers

(See par. 27, TM 12-425.) The value of the interview varies directly with the skill of the interviewer and it is at its best only when conducted by one competent in the use of this tool. Expert interviewers are hard to find but the Army has gone far in locating such personnel and in increasing their proficiency through training. Some of the more important qualifications of good interviewers are listed.

- a. General intelligence well above the average.
- b. A definite interest in people.
- c. A willingness and ability to take the point of view of the other person.
- d. The ability to evaluate and to discount one's own prejudices.
- e. The art of listening as well as the art of conversing.
- f. The art of gaining the confidence of the other person.
- g. The ability to convince the other person of a genuine desire to be helpful.
- h. The ability to exercise tactful control of the interview without a domineering attitude.
- i. A skill in observing and evaluating behavior.
- j. A well-practiced skill in separating fact from inference.
- k. A broad knowledge of civilian and Army occupations and requirements.
- l. A good military bearing and a frank, straightforward manner.

### 96. Conduct of Interview

There is no fixed formula for the interview. It will vary according to its purpose and according to the personalities of the two participants. It must be flexible in form and readily adaptable to personnel requirements and idiosyncrasies. Yet there are several principles that should apply in most instances. Many of these points have already been outlined in connection with the individual testing session (ch. 4); that section should be carefully reviewed. A brief enumeration of the salient features of the successful interview is here given as a further guide.

- a. The *physical setting* of the interview is important. A quiet room, or at least a separate booth, will provide a measure of privacy that is essential.
- b. The *interviewer's preparation* for the interview should include a definition of its general purpose, a schedule of points to be covered, and a thorough review of all information already available concerning the soldier, including all data obtained through tests, questionnaires, and rating scales.
- c. The *orientation of the subject* should start with an effort to create a friendly atmosphere, to put him at his ease, and should include a straightforward statement on the purpose of the interview.
- d. The *tempo* of the interview should be adjusted to the needs of the soldier. He should be given plenty of time to answer in his own way, to amplify and elaborate, without feeling hurried.
- e. The *responses of the person being interviewed and any pertinent observations of the interviewer* should be recorded in some form during the session. Brief notes, symbols, marks on previously prepared check lists—any or all of these methods should be used in order that a complete report of the interview may be prepared.

## CHAPTER 7

### AUTHORIZED ARMY TESTS— USE IN INDUCTION STATIONS, RECEPTION CENTERS, AND OTHER ARMY INSTALLATIONS

#### Section I. PSYCHOLOGICAL TESTING IN INDUCTION STATION

##### 97. General

Since it cannot be expected that every civilian possesses the physical, mental, and moral characteristics essential to the effective soldier, it is necessary to exercise care in the selection of candidates in order to reduce to a minimum the number of failures in training and in service. Laxity in the standards of acceptance would result in the accumulation of unsatisfactory soldier material and increase both the time and money costs of training. It is in the interests of both economy and efficiency that the Army has set certain minimum standards for induction. Moreover, the data gathered during the screening process is an invaluable aid to assignment throughout the career of every soldier. This section deals with the specific tests, methods, and procedures of psychological screening employed in induction stations.

##### 98. Induction Standards

The standards for induction into the armed forces have been specified by the War and Navy Departments. They may be classified into three categories:

- a. Administrative, including age, citizenship, and character.
- b. Medical (physical and neuropsychiatric).
- c. Psychological or intelligence.

Standards in the first category are prescribed by AR 615-500. Physical and neuropsychiatric standards are outlined in MR 1-9. Minimum intelligence standards are defined by section XXII of MR 1-9 in terms of educational achievement and passing scores on authorized "objective tests of intelligence." The tests to be used, and the scores on these tests to be considered as passing, are specified by directive.

##### 99. Procedures at Induction Stations

The procedure is essentially one of screening men according to the three categories of standards listed in paragraph 98. Each requires its own methods

and specially trained personnel. Moreover, each part of this screening process provides data of value to all who will subsequently assign the inductee to any duty or deal with him in any responsible, authorized, and military fashion whatsoever.

##### 100. Psychological Standards

According to paragraph 100, MR 1-9: "Individuals who are graduates of standard English-speaking high schools are acceptable. Individuals who are not graduates of standard English-speaking high schools will be given prescribed objective tests of intelligence. A man achieving the critical score or a higher score on one or more of the authorized tests is acceptable for induction." The psychological screening process designed to implement this regulation is composed of two main parts—a preliminary screening interview and a series or battery of objective tests.

##### 101. Preliminary Screening Interview

It is the primary purpose of this interview to determine whether or not the selectee is a graduate of a "standard English-speaking high school." Consequently, it is a brief routine interview of the fact-finding type. Satisfactory evidence of academic achievement may be obtained from the DSS Form No. 221, from a transcript of the high school record, certificate, diploma, or even from a written or verbal statement. The presentation of such satisfactory evidence of graduation from a "standard English-speaking high school" is a sufficient basis for a recommendation of inductibility, insofar as mental qualifications are concerned. The lack of such evidence, however, may *not* be considered grounds for rejection. No man is rejected for failure to meet the psychological or intelligence standards until he has demonstrated his inability to qualify for induction on the basis of the prescribed psychological tests.

##### 102. Psychological Testing Program

- a. **SELECTEES TESTED.** Only selectees who have not completed the required course of study in a

# PSYCHOLOGICAL SCREENING IN THE INDUCTION STATION

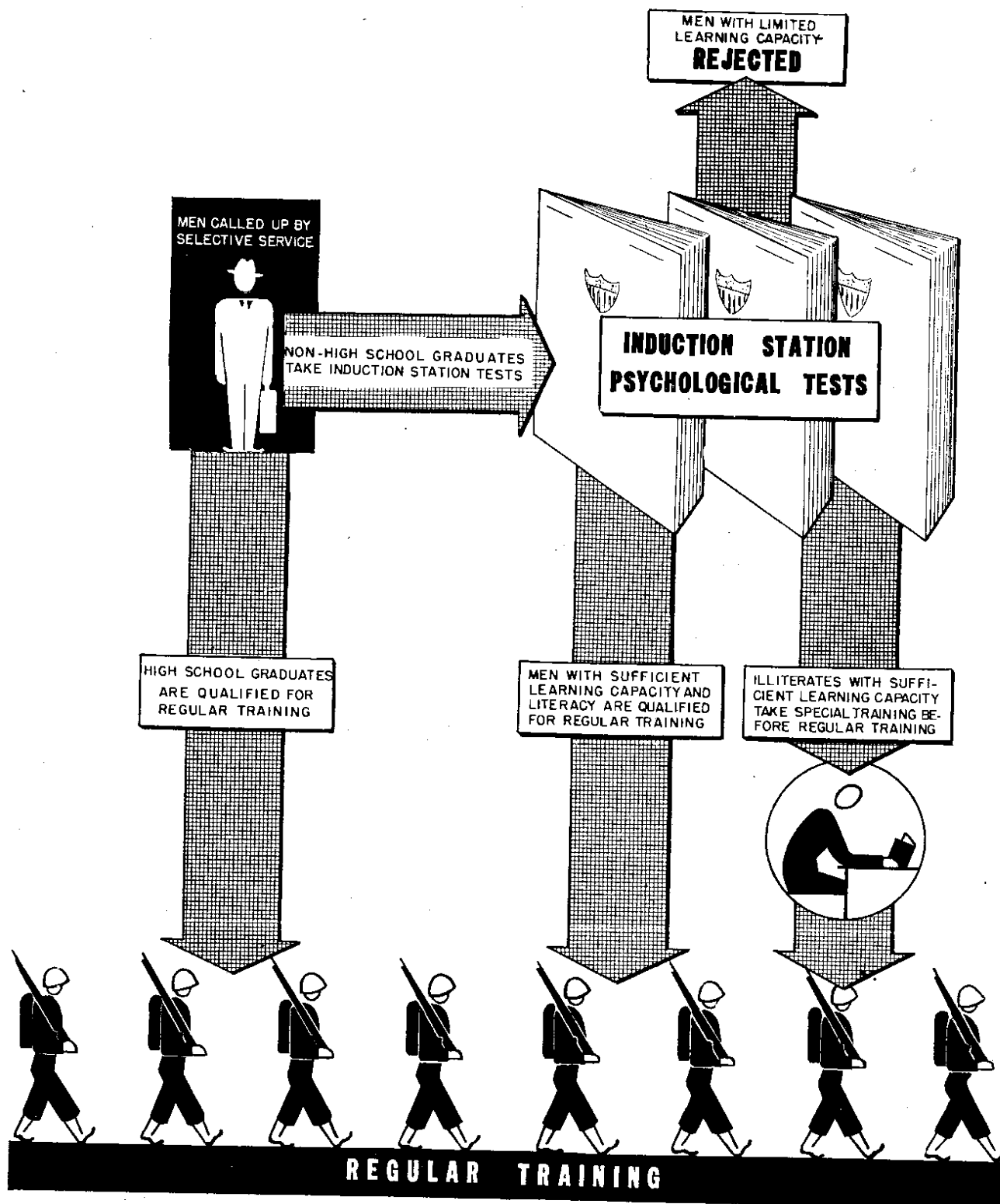


Figure 31.

standard English-speaking high school are referred to the psychological testing section of the induction station. There they are subjected to one or more of a series of tests designed to determine their acceptability for induction. The tests currently authorized for this use are:

- (1) The Qualification Test (Q-1 or Q-2).
- (2) The Group Target Test (GT-1).
- (3) The Individual Examination (IE-1).
- (4) The Nonlanguage Individual Examination (NIE-1).

These tests were designed specifically for induction station use and are restricted to installations of this type.

b. **TESTING PROBLEM.** The fundamental aim of the induction station psychological tests is to insure that non-high-school graduates among selectees accepted for induction possess the minimum mental capacity to absorb military training and become useful or satisfactory soldiers. Non-high-school graduates include men who are literate and have the necessary ability to learn, but whose formal education has been cut short for one reason or another. Also included are men who have gone to school in a foreign land, men who are illiterate because of lack of educational opportunities, and men whose intellectual capacity is seriously limited. Those in the last category would constitute a definite detriment to the Army if inducted. The non-English and illiterate may be useful if they can obtain a sufficient mastery of the simple verbal and numerical skills that are basic to satisfactory completion of military training. Non-high-school graduates thus fall into three general categories:

- (1) Those possessing sufficient capacity to learn and sufficient literacy to undertake basic training.
- (2) Those possessing sufficient capacity to learn, but requiring further literacy training before undertaking basic training.
- (3) Those whose capacity to learn is inadequate.

c. **INDUCTION STATION TESTING BATTERY.** The task of the induction station testing program is to identify and to separate these three groups. Since the men to be classified fall into more than two groups, no single testing instrument can do the job. A battery of tests must be used, each performing a further step in the screening process.

- (1) *Qualification Test (Q-1 and Q-2).* This test is a quick screening device to identify those

men who belong in the first category—that is, men whose capacity to learn and whose literacy level are both sufficient to recommend immediate assignment to basic training. The test consists of 17 carefully selected items of the free-answer type, dealing with such matters as paragraph reading and comprehension, simple arithmetic, and what might be called general intellectual functions (such as the understanding of directions, relations, etc.). Some idea of the difficulty of the items can be gained from the fact that the average score of non-high school graduates is approximately 12. The test may be administered to groups of any convenient size; scoring is relatively simple and rapid. To allow for retesting, two alternate forms of the test are currently authorized, Q-1 and Q-2. They were standardized by administering them to some 3,300 non-high-school graduates in induction stations throughout the country. The distributions of scores for the two forms, or the percentage of either group receiving any given score, are very nearly identical, and for all practical purposes the two forms may be used interchangeably. The Qualification Test differentiates among those men who rank low in general learning ability as compared with all Army men. Since its content is all verbal or numerical and the examinees are required to read, scores on the test closely parallel those obtained with instruments which measure literacy. A very high percentage of those who pass the test are both sufficiently literate and sufficiently equipped with mental capacity to complete basic military training in a satisfactory manner. Consequently, all who pass this test are classified as inductible and qualified for regular training assignment.\*

- (2) *Group Target Test (GT-1).* Qualification Test failures include persons in both the second and third categories, paragraph 102b(2) and (3). Separating them is the job of the remaining tests in the induction station testing program. Since all these men will have been tried (on the Qualification Test) and found wanting in verbal facility, these remaining tests plumb the depths of mental endowment without recourse to printed question or written answer. They are, therefore, performance tests requiring a minimum of verbal ability. The first of these is the Group Target Test. As the name implies, it is a group test. The examiner gives certain simple verbal directions and with a pointer indicates

\*The Qualification Test is one of the few Army tests for which raw scores are not converted to a standard score scale. The specific score which is considered as passing is stipulated by directive, and is subject to change up or down in accordance with the manpower requirements of the armed forces. As of present date, the critical or passing score is 9.



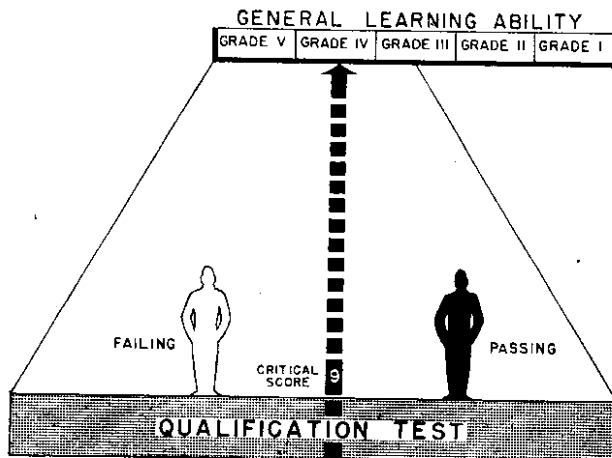


Figure 32. The qualification test differentiates among men who rank low in general learning ability.

certain movements on the wall chart. The examinee gives his responses by drawing lines connecting appropriate dots in the "pictures" on his record sheet. Each of the 28 "pictures" is for a different problem or item. These are of three types: perception and memory of patterns, pattern orientation, and directional orientation. The raw score, or number of correct responses, is converted into a standard score based on the performance of the standardization population—a large representative sampling of men from the lower half of the distribution of Qualification Test scores in induction stations serving widely different geographical and cultural areas of the country. Unlike the usual Army standard score scale (see ch. 5) the scale for induction station tests runs from 0 to 60 with an average of 30 and standard deviation of 10.

(3) *Individual Examination (IE-1)*. For the bulk of the Group Target Test failures, the Individual Examination is the final hurdle. As its name implies, it is an individual test. In view of its position in the program, it requires neither reading nor comprehension of verbal instructions other than the simplest. The test is composed of two main parts. Part I gets at the examinee's ability to adapt his timing and coordination to a specified pattern by six "marching" problems of increasing difficulty. He is required to "march" up pathways of circles and lines, with the two hands alternately, scoring "hits" in the left- and right-hand pathways, while keeping in time with a 120-per-minute cadence counted out by the examiner. The score on each problem is the number of successful steps or "hits" in the *left-hand* pathway before an error is made.

The score for the whole of Part I is the sum of the scores on the six problems. Part II consists of nine brief problems involving such tasks as the reproduction of patterns with blocks, the arrangement of figure sequences, the memory for designs, and simple directions. The score for the whole examination is expressed in standard score form in the same manner as the Group Target Test scores.

(4) *Nonlanguage Individual Examination (NIE-1)*. Men of foreign extraction who are not literate in English may be unable to answer any of the items of the Qualification Test, and if their understanding of spoken English is limited, may also be handicapped on the Group Target Test and the Individual Examination. Such men may be educated in their native tongues, and given the minimum essentials of English verbal facility, may become valuable soldiers. The Nonlanguage Individual Examination is employed with those men who fall in this category. It is administered to men who have scored zero on the Qualification Test and have subsequently failed the Group Target Test, when, in the opinion of the induction station personnel consultant, they may be classified as non-English-speaking. The content of the test is completely nonverbal and is administered by means of pantomime, gestures, and demonstrations. Oral instructions, even in the examinee's native tongue, are not allowed. A typical page of the test contains two series of pictures, figures, or designs, one series at the top of the page and one at the bottom. The examinee merely draws lines on the page connecting like or similar pictures in the two series. The problem, of course, lies in the detection of the resemblance, and this varies from identity (as for example, a tank in each series) through similarity of function to the rather obscure likeness of a pair of designs. The raw score is the number of correct pairings, and this is converted into a standard score in the same manner as the other tests in the battery.

### 103. Recording Scores

The scores received on each of the tests in the induction station series are recorded on the work sheet which accompanies each man through the entire screening process, and on the DSS Form 221. If the individual is accepted for induction the latter form will go with him to the reception center, where the scores will be transcribed to his WD AGO Form 20, giving the official abbreviation of the test and the score. For the Qualification Test, the raw score

is recorded, while for the remaining tests the conversion into standard score form is used, as shown in the example below:

(18)—① Other tests	
Test	Grade—Score
Q-1	7
GT-1	25
IE-1	29
NIE-1	30

#### 104. Summary of Induction Station Test Procedures

a. Men who have not graduated from standard English-speaking high schools are administered the Qualification Test and, according to their performance, are separated into three classes:

- (1) Those whose score is 9 or above.
- (2) Those scoring 1 through 8.
- (3) Those scoring zero.

b. The first class is considered inductible and ready to embark upon regular basic training. The second and third classes are given the Group Target Test. Those in the second class who pass the Group Target Test, or, failing this, pass the Individual Examination, are classified as acceptable for induction, but are qualified for Special Training Unit assignment. There is one modification of this general rule—men in this class who score below 15 (standard score) on the Group Target Test *may* be rejected, at the discretion of the personnel consultant, without attempting the Individual Examination. Those in the third class (Qualification Test scores of zero) who pass the Group Target Test are likewise acceptable for induction and are qualified for Special Training Unit assignment. If a man who understands the English language fails, he is considered noninductible and is rejected following a terminal interview. If a man in this third class fails the Group Target Test, and it is determined that he does not understand English, he will be given an opportunity to demonstrate his ability on the Nonlanguage Individual Examination. A passing score on this test will qualify him for induction and for Special Training Unit assignment, regardless of his performance on the previous tests. A score below passing will be cause for rejection.

#### 105. Validity of Induction Station Tests

The aim of the induction station testing program is to select and qualify for induction those among the non-high-school graduates who are likely to become satisfactory soldiers, and to identify and eliminate those who threaten to become detrimental burdens to an efficient Army team. That the program accomplished this end was effectively demonstrated through the extensive research that led to the development and selection of the battery of tests employed. More men rated on performance in basic infantry training as satisfactory soldiers and fewer rated as unsatisfactory are accepted by the foregoing selection procedure than would be the case if selection were made without tests or on the basis of procedures previously employed for the same purpose. Fewer satisfactory and more unsatisfactory soldiers are rejected.

#### 106. Tests Previously Authorized for Psychological Screening at Induction Stations

a. *The testing program* described above was authorized and instituted as of 1 June 1944. Previous to this time, since September 1942, other tests were employed for the same purpose. Since many men now in the Army were selected by those other tests, and their scores entered into Army records, a brief description of each of these earlier instruments will be given here. They include:

- (1) The Qualification Test;
- (2) The Visual Classification Test.
- (3) The Individual Battery: Wells' Concrete Directions Test and the Block Counting Test.

b. *The Qualification Test* is the same as that currently used and described previously. In the earlier program, however, the critical score was 7. Men reaching or exceeding this score were classified as "literate—inductible," and were consequently qualified for regular assignment. Men scoring 6 or below were classified as "illiterate"; if they passed one of the subsequent tests they were acceptable but qualified for initial Special Training Unit assignment. If they failed all subsequent tests, they were considered non-inductible.

c. *The Visual Classification Test* was designed as a group test of general learning ability requiring no language understanding either to comprehend the instructions or to indicate the responses. Its content was completely pictorial. Each item consisted of five illustrations of objects; four of these belonged to the same class or category, the fifth

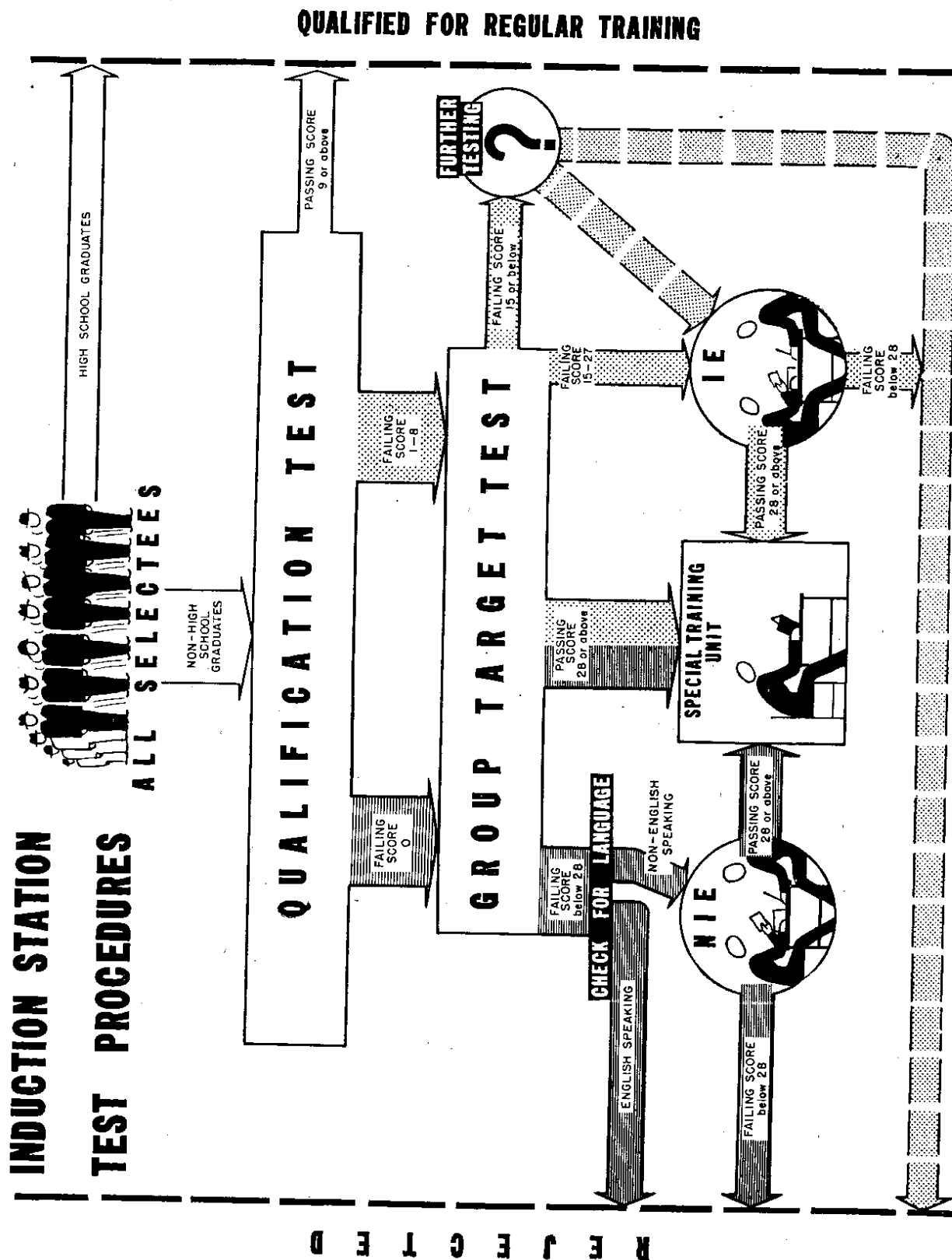


Figure 33.

being different. By means of pantomime and simple oral directions, the examinee was directed to identify and cross out that picture in each series which did not logically belong in the same category with the remaining four.

d. *Individual Tests* were designed for men who failed both the Qualification Test and the Visual Classification Test. These men were put through a final screening with the two individual tests—the Wells' Concrete Directions Test and the Block Counting Test. Both of these tests were of the performance type, the former involving the relational manipulation of some common tools and objects, the latter requiring the correct count of the blocks contained in a pictured aggregation.

e. Table II summarizes some of the pertinent data on these earlier tests. They are presented here as an aid to the evaluation and interpretation of scores which will be frequently found in the records of men throughout the Army.

### 107. Terminal Interview

Before any selectee is rejected for failure to meet the minimum psychological standards for induction, he is given a final brief interview. During this interview an attempt is made to discover and account for any discrepancies between test performance and previous personal history. Low test scores are inconsistent with adequate educational or occupational experience and may suggest faulty test administration, misinterpretation of instructions, lack of proper motivation, or deliberate malingering. Even clerical errors committed in scoring, adding, or recording results may on occasion account for startling inconsistencies. Whenever the interviewer has reason to believe that the test score does not accurately reflect the selectee's

true ability, he may recommend a reexamination on any of the tests in the series. Authorization to re-test should never be abused, however. *The purpose of reexamination is not to boost induction rates at the expense of Army efficiency, but to insure the precision of selection procedure.*

## Section II. PSYCHOLOGICAL TESTING IN RECEPTION CENTER

### 108. General

When the selectee has been examined and found to be a suitable prospect for military service, he moves on to the reception center for equipment, orientation, classification, and assignment. Personal history data are collected and systematized; mental abilities and aptitudes are determined. All of this information is recorded, coded, and punched on the WD AGO Form 20, Soldier's Qualification Card. This card accompanies each man throughout his military career and serves as a continuous and cumulative record of his experience, training, and qualifications.

### 109. Classification Interview

One of the most important phases of reception center procedure is the individual interviewing of each incoming soldier. It is during this interview that the WD AGO Form 20 is initiated, and in view of the emphasis placed upon this document in all selections and assignments affecting the soldier throughout his Army service, it is evident that the interview is of vital importance to the Army and to the soldier himself.

a. **SOLDIER'S QUALIFICATION CARD.** The WD AGO Form 20 (Soldier's Qualification Card) is the basic instrument of the Army classification system.

Table II

Important Statistics of Earlier Induction Station Tests			
Name of Test	Official Abbreviation	Maximum Score	Passing Score
1. Qualification Test	Q-1 or Q-2	17	7
2. Visual Classification Test	VC-1a	50	36
3. Wells' Concrete Directions*	CD	63	52
4. Block Counting Test*	DST-10	16	12

\*Passing scores on both individual tests were required for acceptance.

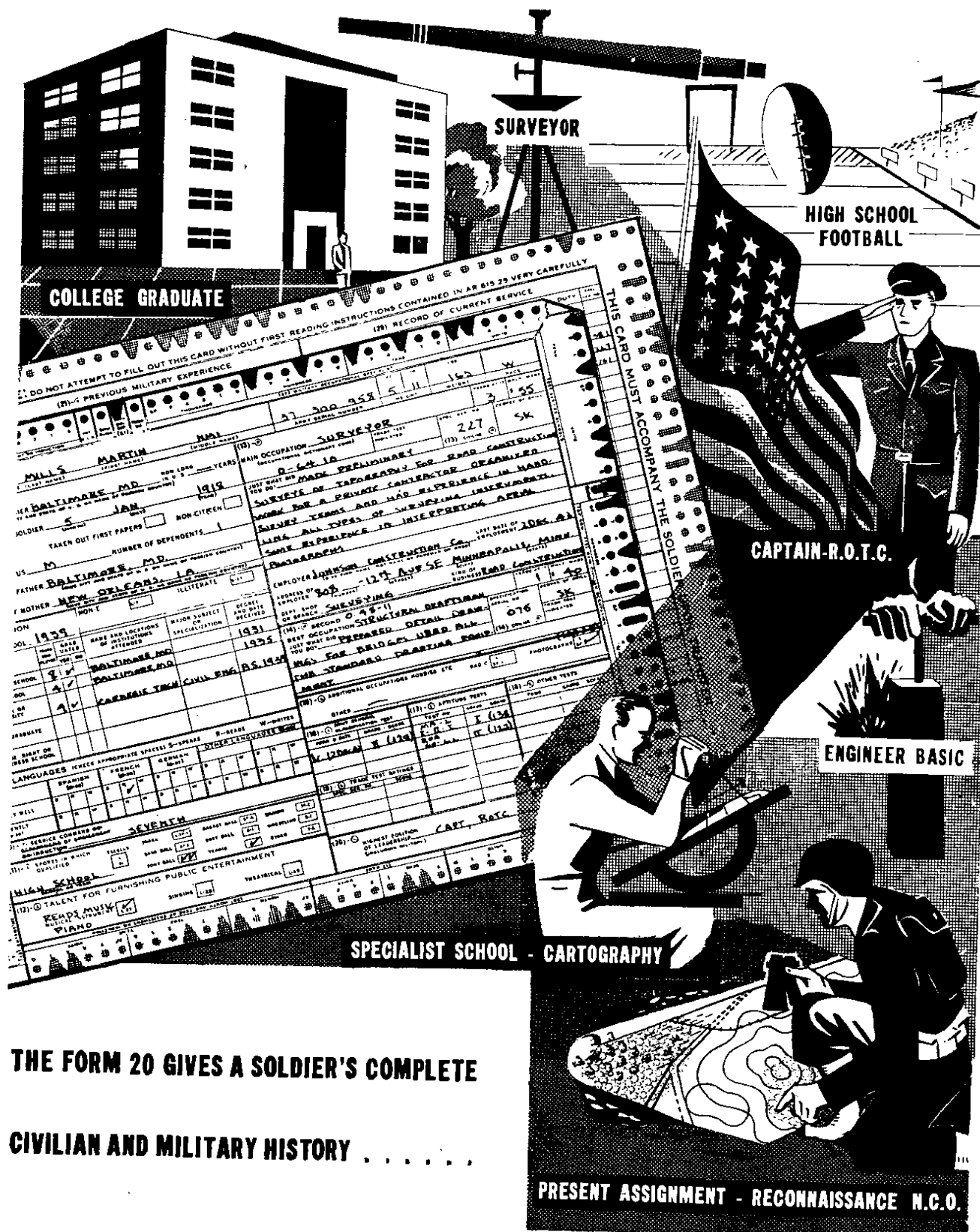


Figure 34.

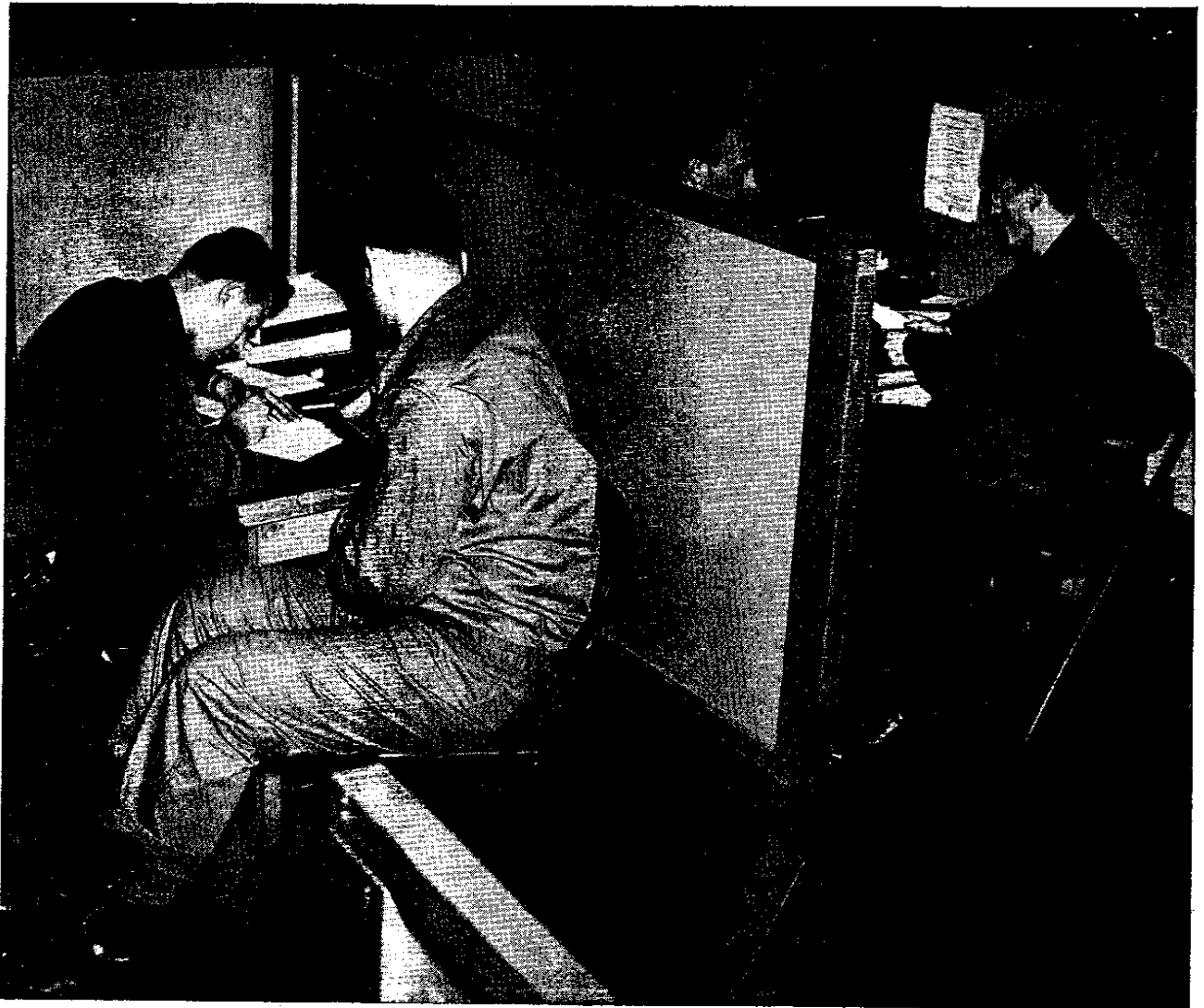
It is designed to provide spaces for all information about the soldier that may be helpful in evaluating his past experience and in judging his potentialities for future development. It may be said that the most valuable resources of an Army are the skills, potentialities, and training experiences represented in the men who make up that Army. With the number of these men running into the millions, the value of the Soldier's Qualification Card becomes more evident. When a group is small, it is possible for those in command to become intimately acquainted with each of the men—to learn at first hand what the individual has done in the past, how well he executes his assignments, what particular skills he possesses, his capacity for leadership, and so on. If such a group remained the same, one could expect that proper selections and assignments would be made and all men used effectively even though the only personnel records were filed away in the mind of the commander. But the Army is not like that. Transfers at all stages and casualties in combat areas create a condition of continuous change among both officers and men. Replacements are often made from much larger groups where individual differences cannot be clearly observed. Every able officer should know his men, but when he leaves his command, he should pass that knowledge on to his successor. Otherwise the time and effort he has expended in training his men and in cataloguing their qualifications and skills may be largely wasted. His knowledge that Corporal Smith is an efficient typist, or that Private Jones has learned, under his tutelage, the secrets of scouting and patrolling may give him some satisfaction. But unless this information is recorded and passed on, Corporal Smith's new commander is quite likely to overlook him when picking a new company clerk, and Private Jones may be trained all over again for a different assignment. The WD AGO Form 20 is a record of experience and training which transfers knowledge of men. Through its proper use, any officer can know his men and employ them effectively even before seeing them. Corporal Smith's card, for instance, will show his civilian experience as a typist, his degree of skill, his performance on a typing test, and any previous experience as company clerk. Some of this information will be coded and punched in the margins of the card so that Corporal Smith, along with every other typist in the group, can be identified in a few minutes time. In a sense, the card is a talking picture of the men,

showing what he is like and telling what he can do. Its value will depend on the extent to which it is an accurate and faithful reproduction. Its usefulness will depend upon the intelligence with which responsible officers apply the information the card supplies.

b. **PURPOSE OF THIS INTERVIEW.** The purpose of the classification interview is to discover and enter upon the Soldier's Qualification Card all information about the man that will go toward making that record the most complete and accurate picture possible at this early stage of the soldier's career. Such information will include his age, physical condition (Physical Profile Serial, reference, MR 1-9 Supplement, 30 June 1945), education, foreign language proficiency, civilian occupational history, hobbies, previous military experience (ROTC, CMTC, CCC, etc.) and positions of leadership. It will also include scores he has made on certain authorized tests that are administered to all new soldiers. A detailed discussion of each of the items of the WD AGO Form 20 is presented in TM 12-425, which should be used at all times as a guide to the interviewing procedure.

c. **IMPORTANCE OF THIS INTERVIEW.** A more general explanation of the nature of the interview and the selection and training of interviewers is contained in chapter 6 of the present manual. The material presented there should be carefully studied. The use to which the information will later be put should serve as a constant reminder of the importance of good interviewing. Some soldiers are not sufficiently impressed with the importance of the interview. They look upon it as a tedious but necessary routine, and consequently may not tell about some minor but valuable work experience. Some seem to feel that it is good Army practice to speak only when spoken to—to give only "yes" or "no" answers to direct questions. Others will embroider the most trivial facts into high-sounding statements intended to impress. The successful interviewer will create the proper impression to open the conversational floodgates. And he will seine the stream of talk, throwing back the small ones and saving the larger ones for the record.

d. **ORAL TRADE QUESTIONS** (developed by United States Employment Service). The most significant and important single item of the WD AGO Form 20 is that dealing with the soldier's main civilian occupation (MCO). If the Army is to make use of all important civilian occupational experiences and skills of its men, it is essential that such occupa-



## THE INITIAL CLASSIFICATION INTERVIEW AT THE RECEPTION CENTER

Figure 35.

tional experience must be properly classified and coded, and that some indication of the soldier's proficiency in that field be recorded. Unfortunately, it is not always possible to obtain such information by the simple procedure of asking the man. Because of the lack of uniformity in occupational titles, the differences in standards of proficiency, and, sometimes, the plain "cussedness" of human nature, the man's own statement of his civilian job and his skill in it will not lead to accurate classification. The man who lays claim to the trade of bricklayer may, in reality, have been a hod carrier, a bricklayer's helper, or a carpenter. Or he may

have specialized in a narrow phase of the craft and be inexpert in other essentials. Consequently, it is desirable to check this man's knowledge with reference to the trade or occupation in which he claims to be skilled.

(1) The Oral Trade Questions are standardized sets of questions relating to a variety of skilled trades. In each set, the questions are specific and require specific answers; the wording is simple and in the language of the worker. Each set contains approximately 15 questions carefully selected on the basis of their actual discrimination between groups of workers of known characteristics. Correct

or acceptable answers are all contained in the manual, and the score is simply the number of questions answered correctly. The extent to which they differentiate degrees of skill is illustrated in table III which shows the average (median) scores for three groups of workers on the questions for bricklayers.

Table III

Average Score for Three Groups of Workers on the Oral Trade Questions for Bricklayers

Occupational group	Average score
Expert Bricklayers	12
Bricklayer Apprentices and Helpers	5
Workers in Related Fields	1

(2) In use, the Oral Trade Questions become a part of the classification interview. They should be prefaced by some such remark as "I'd like to ask you a few questions about the work you did before coming into the Army." They should be read in a natural conversational tone, but they should never be altered or enlarged upon in any way. Scores are interpreted according to the norms accompanying each set of questions in the manual, and they should be recorded on the WD, AGO Form 20 in the manner specified by TM 12-425.

### 110. Reception Center Testing Program

Among the items of information that are recorded on the Soldier's Qualification Card are the scores on all authorized tests which the soldier may take during his Army career. The first are scores of his induction station tests. The next are scores of tests given to all soldiers entering reception centers. All of these have widespread application throughout the soldier's career and should, therefore, be administered as early as possible. Some of the scores of these tests are used immediately in making the initial assignments of the men. There is one exception to this general rule of testing all incoming men: those who have been accepted at the induction station and classified as "illiterate," or qualified for Special Training Unit assignment, will not be given the reception center tests until they have satisfactorily completed this course of special training in reading and arithmetic.

\*Four noninformation tests are available at present. Other tests measuring knowledge of automotive mechanics and general shop work are still in the process of development.

### 111. Army General Classification Test

a. The first tests to be administered in the reception center series are those which constitute the Army General Classification Test-3 (AGCT-3). As the name indicates, these tests measure abilities and aptitudes which underlie a large number of Army assignments. They are given at the reception center in order to obtain an early estimate of the assignment or training regime in which the soldier is likely to be most proficient. There are four tests in the battery,\* each yielding a separate score of a separate important skill. A general over-all score is obtained by summation of these four separate scores. Performance on each of the four separate tests and total test performance are expressed in standard score form and in Army grades. (See ch. 5.) The total scores are comparable to those obtained with the previous editions of the Army General Classification Test (see par. 113a) which have been administered to over eight million men and which have found widespread and valuable employment in the selection of men for all kinds of Army specialist training and assignment. Thus, the total AGCT-3 score may be interpreted as a measure of general learning ability, or the factor commonly referred to as "general intelligence."

b. The major difference between the present and previous forms of the Army General Classification Test is the provision in the AGCT-3 for separate scores for each of the abilities or aptitudes measured in addition to the total over-all score. This difference also constitutes the chief advantage of the present form of the test. While it is undoubtedly true that the composite measure of general learning ability is basic to success in most Army training, Army training courses and assignments differ materially in the demands they make upon particular skills. The company clerk, the automotive mechanic, and the fire control computer will all require a given measure of general intellectual capacity. But this general capacity, as the name implies, is compounded of a variety of specific skills. The man who scores high on tests of "general intelligence" is the one who knows a lot about a variety of topics. Yet he will usually be more skilled or better informed about some of these topics than others, and these peaks and valleys in his mental makeup should be taken into account in selecting him for a particular assignment. If



his verbal or language facility is more highly developed than his mechanical or mathematical facility, he should qualify for such assignments as clerk or instructor which place a premium on verbal ability. On the other hand, surveying and controlling the fire of coastal batteries often require a mathematical ability of a higher order. It is, then, the primary advantage of the AGCT-3 that it provides reliable and thorough measures of some of these specific abilities of which general ability to learn is composed.

c. The four tests of the AGCT-3 are: Reading and Vocabulary; Arithmetic Computation; Arithmetic Reasoning; and Pattern Analysis. The nature and content of the first three of these is self-evident, and the usefulness of the abilities they measure (verbal and numerical) in predicting success in training is well established. The Pattern Analysis Test is an improved version of part II of the Mechanical Aptitude Test earlier included in the reception center testing program. It is composed of line-drawn pictures of three dimensional objects accompanied by outline patterns from which these objects are formed. Various edges of the picture and lines of the pattern are marked, and it is the examinee's problem to match up corresponding edges and lines. In other words, the test measures skill in the mental manipulation of spatial relations and the visualization of three-dimensional form. That such skills are related to or predictive of mechanical proficiency in the Army has been well established. One illustration will suffice. The present Pattern Analysis Test was administered to airplane mechanic students at the AAFTC, Keesler Field, Mississippi, and test scores compared with grades in the course. The extent to which scores on the test can be used to predict course grades is illustrated in the following table:

Table IV

Relation Between Pattern Analysis Test Score and Grades in Airplane Mechanics Courses  
(AAFTC, Keesler Field, Miss.)

Men receiving a test score of:	100	110	120
Have these chances in 100 of reaching or exceeding the average grade in course:	55	63	75

From this table it is evident that the predictive value of the Pattern Analysis Test is high; men scoring 120 on the test, for example, stand a 3 to 1 chance of being average or better students in the course.

d. The Army General Classification Test-3 constitutes a comprehensive and flexible set of measuring tools. Taken all together, they provide an accurate and valuable indication of general ability to learn in a wide variety of Army situations. Taken singly and in combination, they can serve as useful predictors of success in specific courses of training or specific assignments. The composite score finds immediate use in the assignment of men to the training centers of the various arms and services, and later, to service and combat units, since the advisability of building balanced units has been abundantly demonstrated. The specific scores, considered in conjunction with this over-all score, are most advantageously employed in the selection of men for specialist training.

## 112. Army Radio Code Aptitude Test (ARC-1)

One of the outstanding features of modern warfare is the extensive employment in all echelons of all arms and services of the most advanced technological improvements in such fields as transportation, engineering, ordnance, medicine, and communication. A mobile Army, to maintain effective control, must be coordinated by a complicated network of communications. More than ever before, radio is playing a major role in the conduct of hostilities. Most of this is voice radio, to be sure; nevertheless, the Army's needs for radio code operators are larger than can be met by the selection of all experienced code operators who are inducted into the Army. Consequently, the training of radio code operators is an urgent necessity. Experience has demonstrated marked individual differences in aptitude for learning to receive code messages, and research has shown that this aptitude can be predetermined by means of suitable tests. Such a test is the Army Radio Code Aptitude Test (ARC-1), given to all men at reception centers for the purpose of predicting probable success in this kind of training.

a. The Army Radio Code Aptitude Test (ARC-1) is presented by means of phonograph recordings. It involves the learning and recognition of the code signals for three specified letters of the alphabet. The test proper is preceded by a learning period during which these code signals are sounded accompanied by an announcement of the corresponding letter. In the test itself, the signals alone are presented in random order, and the examinee must indicate which of the letters the signal represents. The first half of the items are presented at a speed equiv-

alent to approximately 11 words per minute, and the last half at approximately 15 words per minute. The score for the test is the number of correct recognitions minus one-half the number wrong, and this is converted into standard score terms in the usual manner.

b. That the Army Radio Code Aptitude Test is capable of predicting probable success in training to receive code signals has been amply demonstrated. The following table (table V) illustrates the relationship between test scores and code-receiving speed in several classes after two months of training. This relationship is expressed in terms of the percentage of men within certain score ranges on the test who had reached or exceeded a code-receiving speed of 12 words per minute (the average rate for this stage of training). Thus, of men scoring 130 or above on the test, 87 percent had reached or exceeded the average rate, after two months of training, of 12 words per minute. Test scores within this range are, therefore, considered as superior, and anyone scoring this high stands a chance of better than 6 to 1 of making a satisfactory completion of the course of training. On the other hand, persons scoring below 100 (and since 100 is the average standard score, this means half of the Army) stand about the same chance (6 to 1) of poorer-than-average performance in the course.

### 113. Earlier Reception Center Tests

The tests so far described as constituting the reception center testing program were developed after months of research and experiment. But while these were being developed, other tests were being used, and their results will be found in the records of millions of men now in the Army.

a. **ARMY GENERAL CLASSIFICATION TEST-1.** Perhaps the most widely known of all Army tests is the Army General Classification Test-1 or AGCT-1. In all, four forms of the test were developed. The first two forms, AGCT-1a and 1b, after serv-

ing with approximately four million men, were superseded by AGCT-1c and 1d. These editions in turn were used with the next four or five million men. Scores were widely used. They played a prominent part in certain assignments. They were used in the selection of men for specialist schools and for officer candidate training. Consequently, the test enjoyed wide publicity—and suffered extensive misinterpretation, especially since it was widely—though wrongly—known as the “IQ Test.” The AGCT-1 was composed of three kinds of questions—vocabulary, arithmetic, and “box counting.” Its major disadvantage was its inability to yield separate scores of these three abilities. Nevertheless, it was of undoubted value as a thorough and reliable measure of general learning ability, and, as subsequent chapters will show, it produced results which were directly related to success in many different kinds of courses of training.

b. **MECHANICAL APTITUDE TEST (MA).** This test was designed to estimate chances for success in training for assignments of a mechanical nature. The test was in three parts, each scored separately. The part scores were converted and recorded as Army grades while the total score was expressed as a standard score as well as an Army grade. In the first form of the test (MA-1) the three parts were: mechanical movements; surface development; and shop mathematics. This form was superseded by MA-2 and MA-3, each of which was composed of the following parts: mechanical information; surface development; and mechanical comprehension. (A fourth form, MA-4, was designed and authorized for use with the WAC.) In general, high scores on the test are indicative of an aptitude for training in technical or mechanical courses such as those for motor mechanics, artillery mechanics, or aircraft armorers.

c. **RADIOTELEGRAPH OPERATOR APTITUDE TEST (ROA-1, X-1).** This was the test used earlier to aid in the selection of radio code operators. A man

Table V

Performance in Code-Receiving Speed After Two Months of Training Compared with Scores on the ARC-1		
Interpretation	Standard score range	Percent receiving at the rate of 12 words per minute or better
Superior	130 and above	87
Satisfactory	110-129	70
Low	100-109	48
Unsatisfactory	Below 100	32

who is able to differentiate code patterns which he hears will, in general, learn code more quickly than will a man who does not recognize such differences. He will, therefore, be a better risk for training as a radio code operator. The Radiotelegraph Operator Aptitude Test consisted of two consecutive administrations of the old Signal Corps code aptitude test. It was presented by means of phonographic recordings. Each item consisted of two code patterns sounded in succession, and the examinee was required to decide whether the two patterns were the same or different. In general, men who have had previous experience with code, or men who play a musical instrument, tend to differentiate better and to make higher scores, and with such men, the predictive value of the test is apt to be high. With inexperienced men, where the real selection problem lies, the ROA will select fewer men than the ARC-1, and fewer of those selected will become satisfactory operators. Hence, its replacement by the improved test.

### Section III. PSYCHOLOGICAL TESTING IN OTHER INSTALLATIONS

#### 114. Continuing Need for Testing

The use of psychological tests is not limited to induction stations and reception centers but occurs in installations of all types and for many reasons. Personnel selection and initial assignment for training obviously require the kinds of information about a soldier's abilities which these tests furnish; but it is equally true that such information is valuable whenever decisions must be reached regarding changes in assignment or selection for more advanced training. It is for these reasons that each soldier's test scores are recorded along with his personal history data on WD AGO Form 20 (Soldier's Qualification Card) which follows him wherever he goes until he is separated from the service. Although every effort is made to safeguard personnel records there are instances both in the zone of interior and in theaters of operation when it is necessary to reconstruct records and to replace those which have been lost. Likewise, in the transfer of personnel, records are often delayed in transit with the result that information pertaining to test results is not available at the time it is most needed.

In these instances it is necessary to retest these persons in order that test scores may be available. Such testing may involve only the general classification test or it may include any number of specialized tests which are described in other sections of this manual.

a. The need for test information is likely to be acute when soldiers arrive at a training center, redistribution center or a replacement depot. The large assortment of abilities represented must be distributed at once in such a manner that each requisitioning unit receives its due proportion of the available abilities and skills. Likewise the maximum use must be made of the abilities of each man who is available for reassignment. Replacement units in overseas theaters make frequent use of the tests described elsewhere in this manual to aid in the proper assignment of hospital evacuees and other personnel received for retraining and reassignment.

b. Sometimes there are reasons to believe that a man's test score does not correctly represent his abilities. In order to verify or correct the recorded score, the soldier may be given the chance to show what he can do on an alternative form of the test. The regulations pertaining to officer candidates (AR 625-5) stipulate that "a score of 110 or higher in the Army General Classification Test is required." Since other regulations permit a retest of individuals with the AGCT when there is reason to believe previous scores are invalid, a readministration of the AGCT is sometimes warranted when an applicant for Officer Candidate School has demonstrated ability beyond that which is implied by the score on an early administration of the test. In all cases of retesting the provisions of paragraph 18, TM 12-425, will be followed.

c. Clinical diagnosis frequently requires the use of tests in hospitals, disciplinary barracks, consultation sections in training centers, and processing centers. The results of these tests provide valuable aid to clinical psychologists and psychiatrists who are engaged in curative and preventive programs of psychotherapy. Reconditioning and retraining analyses of skills and abilities and, in many instances, personnel measurements are used to assist in determining proper levels and types of placement for convalescing patients. These tests so employed are described elsewhere in detail.

## CHAPTER 8

# AUTHORIZED ARMY TESTS—USE IN SELECTION AND ASSIGNMENT

### Section I. ORIENTATION

#### 115. General

The testing instruments that have been developed for use in connection with the mobilization and general classification of men in the Army have been discussed in previous chapters. The nature of the information obtained by means of these instruments will be clarified in this chapter with particular reference to the usefulness of the data in selecting men for specialist training and for assignment to the particular jobs specified in the tables of organization of the various units. Upon the completion of basic military training, each man in the Army is assigned to a unit, and eventually to a particular job in that unit. All of these jobs have been analyzed and the duties and requirements described in TM 12-427. For many of them, no special preparation is necessary beyond that obtained in basic and unit training, in field exercises and maneuvers. For a large number of jobs, however, special training in the training centers and schools of the various arms and services is required before the duties can be carried out effectively. It is with this latter group, for the most part, that the problems of selection and assignment have been most urgent.

#### 116. Selection for Training

When men are to be trained as specialists, it is obviously essential to insure that only those men are selected for this training who have a reasonable chance of completing the course satisfactorily. If a large pool of qualified men is available, the selection problem is simplified. Previous experience in related fields, education, information, and interests, supplemented by tests of present skills, will reveal those most likely to be satisfactory. When few men clearly qualified on these bases are available, *aptitude tests* must be used to identify the promising men who have no civilian experience in related activities. A number of such aptitude tests have been developed and used in the selection of men for specialist training. They are especially important in evaluating the potential usefulness of younger men who have had little or no job experience previous to induction.

#### 117. Selection for Direct Assignment or Reassignment

The progress of any war is accompanied by changing requirements in techniques, matériel, and men skilled in their use. Victory stems from ability to adjust rapidly to these changes requiring shifting emphases on the various arms and services. Some specialties are reduced in importance; the need for others is augmented. Consequently, men must be reclassified, often retrained, and reassigned. In addition, large numbers of men are returned from the active theaters as casualties to be redistributed and, in many cases, to be retrained. In such cases, the evaluation of military skills acquired through previous Army experience can save much time in selection and avoid much unnecessary retraining. Often it will be a matter of checking whether a man really knows the particular job in which he claims experience or in which he has been classified. At other times, it will be a problem of determining whether he has, by accident or by design, acquired sufficient skill to bypass all or part of the training for a particular job. *Achievement tests* which evaluate the present knowledge of the individual relative to the requirements of the job have been developed for a number of the numerically most important military occupational specialties (MOS). (See current issues of FM 21-6.)

#### 118. Value of Tests for Selection

Tests are employed to select men for training or for particular assignments on the basis of a demonstrated relationship between performance on the tests and subsequent performance in training and/or on the job. The value of any test, then, will be a function of the extent of this relationship. Each Army test is constructed for a specific purpose, and is a valid instrument when used for that purpose. (See chs. 3 and 5.) Some tests are valuable for selecting men for training in particular specialties. Others, like the Army General Classification Test, are useful for predicting success in several kinds of training assignments in which general verbal and numerical skills are involved. Such tests have different validities for each of the various assignments. Where the course calls for considerable exercise of

verbal and numerical abilities, the relationship between performance on the AGCT and success in the course will be high. Where these same abilities are of limited importance, this relationship will be low, and the test of little value in selection for training.

### 119. Expectancy Charts

The extent of this relationship between test performance and subsequent success in training (i. e., the validity of the test for the specific purpose), can be illustrated by means of a chart showing what chances (in 100) a man making a particular score on a given test has of making an average-or-better grade in a particular training course. A number of such expectancy charts are presented in the remaining sections of this chapter. By reference to these charts, the classification officer is able to estimate the probabilities for success of a trainee with a given test score. According to the chart in paragraph 121a(1), for example, a soldier receiving a score of 140 on the Clerical Aptitude Test would have 84 chances in 100 of receiving an average-or-better grade in the clerical course. Or, of 100 such men receiving a score of 140 on the test, 84 can be expected to achieve average-or-better grades in the course. This, in general, is the meaning conveyed by the expectancy chart. There are several limiting factors, however, that must be considered in any correct interpretation of these figures.

a. **ABILITY OF THOSE ASSIGNED TO TRAINING COURSE.** Since the predictions of success in training state the probabilities of reaching or exceeding the *average of the class*, it follows that these probabilities will vary according to the ability level of the class. Where men are preselected for a course on the basis of education or test performance, it will usually follow that the general level of the class is higher. The average (mean) test score of the group studied is given in each of the charts. According to the figures in paragraph 121a(2), for example, the average Army General Classification Test score for the 2,947 clerical trainees studied was 121.7, which is well above the average for the Army as a whole (100). Consequently, one would have to score higher than 121.7 on the test in order to have a better-than-even chance for better-than-average grades in the course. As this chart shows, men receiving a score of 140 on the Army General Classification Test have 76 chances in 100 of achieving average-or-better grades in a clerical course in which the mean AGCT score of the class is 121.7. It

should be noted that each chart includes, in addition to the number of cases involved in the study (N) and the average test score of the group (Mean Score), also the standard deviation of test scores (S. D.) and the coefficient of correlation between test scores and course grades (r).

b. **CONTENT OF THE PARTICULAR COURSE.** Since the expectancy charts are based on a demonstrated relationship between test scores and grades in the particular training course under discussion, it follows that they will continue to serve as guides to prediction only so long as the content of the course remains comparable to that for which this demonstrated relationship was discovered. Where it is known that the content of the course has undergone radical revision, the charts should be used with caution. It is further worth noting, in this respect, that the predictions embodied in the charts often relate to achievement in classroom work and not to those other factors, such as leadership, which may contribute to a successful conclusion of the course.

c. **STABILITY OF STANDARDS.** Availability of men for particular types of training varies from time to time, and course standards are prone to adjust themselves to these changing demands for men. When many more men are available for training than are required as trained specialists, course standards are apt to become more rigorous. The opposite effect results from a decrease in the availability-requirement ratio. Such shifts naturally affect the probabilities contained in the expectancy charts. As standards become more rigorous, the probabilities of success for any given test score becomes less than the figure given in the expectancy chart. Common sense suggests, therefore, that as supply-demand ratios for specialists change, the critical test scores required for selection be adjusted accordingly.

### 120. Plan of This Chapter

The remaining sections of this chapter list tests and present expectancy tables, where available, for many of the areas of training in the Army. For the most part, the various tests are considered within a framework of military occupational specialties (MOS) for which the test has been shown to be a useful instrument. Section II, for example, is concerned with assignments to training for jobs of administrative and clerical nature. It deals with a number of Army jobs in this area, considering for each the tests that are useful in selection for training and those that are designed to implement se-



military occupational specialty who require refresher training before reassignment. Critical scores and interpretations are given in Manual TC-M12 and supplements.

## 122. Clerk-Typist (MOS-405)

a. SELECTION FOR TRAINING. (1) Clerical Aptitude Test.

(2) Army General Classification Test.

b. SELECTION FOR DIRECT ASSIGNMENT. (1) Clerical Experience Check List (TC-23ar).

(2) Clerical Achievement Test (TC-24ar).

(3) Typing Tests are standard paragraphs used to determine the individual's typing proficiency. Results are obtained in "net words per minute," that is, speed minus errors, and should be interpreted in terms of the standards of speed and accuracy required for the particular job for which assignment is considered.

## 123. Stenographer (MOS-213)

a. SELECTION FOR TRAINING. (1) Clerical Aptitude Test.

(2) Army General Classification Test.

b. SELECTION FOR DIRECT ASSIGNMENT. (1) Clerical Experience Check List (TC-23ar).

(2) Clerical Achievement Test (TC-24ar).

(3) Typing Tests (par. 122 b(3)).

(4) Dictation Tests are standard paragraphs used to determine the individual's proficiency in taking and transcribing shorthand. Scores of 65 and above indicate satisfactory proficiency at the rate of 80 words per minute. Scores below 65 indicate unsatisfactory performance at that speed.

(5) Other factors of value in predicting successful performance as stenographer are education and previous experience. Commercial school graduates are likely to be better prepared in the technical phases of the assignment—i. e., grammar and punctuation of correspondence. Men with previous military experience in this field will probably be more acquainted with the forms of military correspondence than those whose training has been chiefly civilian.

## 124. Supply Clerk (Quartermaster Supplies) (MOS-835)

a. SELECTION FOR TRAINING. (1) Clerical Aptitude Test.

(2) Army General Classification Test.

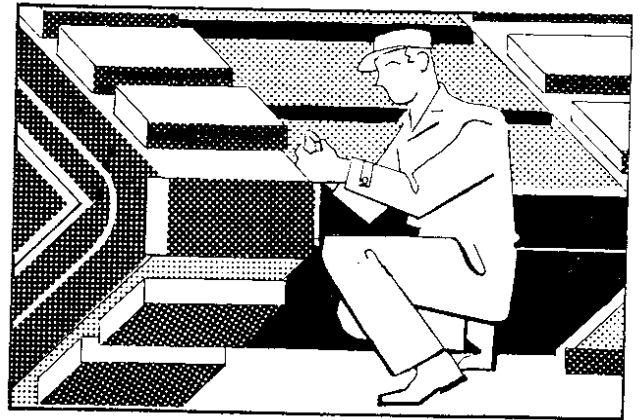
b. SELECTION FOR DIRECT ASSIGNMENT.

(1) The Supply Clerk Experience Check List (Quartermaster Supplies) (TC-29ar) is a con-

venient instrument for securing a general appraisal of a man's experience and background in supply and warehousing clerical work. The number and types of items checked by the average graduate of the Supply Clerk's training course are indicated in a supplement to Manual, Army Trade Screening Tests, TC-M12.

(2) Supply Clerk Test (Quartermaster Supplies) (TC-30ar) is designed to measure technical knowledge of the clerical phases of supply and warehousing. Scores on the test may be used to identify soldiers in this military occupational specialty who require refresher training before reassignment. Critical scores and interpretations are given in Manual TC-M12 and supplements.

## Section III. AUTOMOTIVE AND MECHANICAL ASSIGNMENTS



## 125. Truck Drive, Light (MOS-345) and Truck Driver, Heavy (MOS-931)

SELECTION FOR DIRECT ASSIGNMENT:

a. The Truck Driver Experience Check List (TC-21ar) is designed to furnish a quick and objective estimate of a man's experience in truck driving. The number and types of items checked by the average graduate of a motor vehicle operation course are contained in the supplement to the Manual TC-M12.

b. The Truck Driver Test (TC-22ar) measures technical knowledge relating to the operation of Army motor vehicles. Scores on the test may be used to identify soldiers in this military occupational specialty who require refresher training before reassignment. Critical scores and interpretations are given in Manual TC-M12 and supplements.

126. Mechanic, Automotive (Second Echelon) (MOS-014); Mechanic, Automotive, Wheel Vehicle (Third Echelon) (MOS-965); and Mechanic, Automotive, Track Vehicle (Third Echelon) (MOS-966)

a. SELECTION FOR TRAINING. Mechanical Aptitude Test.

SCORE																	
140									90								
120									71								
100									42								
80									17								
60									5								
<table><tr><td>N</td><td>Mean Score</td><td>SD</td><td>r</td></tr><tr><td>330</td><td>105.0</td><td>17.2</td><td>.54</td></tr></table>										N	Mean Score	SD	r	330	105.0	17.2	.54
N	Mean Score	SD	r														
330	105.0	17.2	.54														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

b. TESTING FOR PLACEMENT IN COURSE. The Army Automotive Screening Test Battery was designed for use in placing men in the course E-70 (Wheeled Vehicle Automotive Mechanics) at the Ordnance School. This course is composed of 3 phases of 4 weeks each, and is basic to all further specialist courses in automotive mechanics. Phase I deals with 1st echelon maintenance, and can be omitted, with a profitable saving in training time and facilities, by men with adequate previous civilian or military experience in automotive maintenance. The Army Automotive Screening Test Battery is used to identify such men. There are 5 tests in the battery.

(1) The Auto Mechanic Experience Check List (TC-13a) is used as an objective means of evaluating past experience in auto mechanics' work.

(2) Tool Usage Film Strip Test (TC-12a).

SCORE																	
130									87								
115									76								
100									61								
85									45								
70									29								
<table><tr><th>N</th><th>Mean Score</th><th>SD</th><th>r</th></tr><tr><td>124</td><td>96.9</td><td>21.7</td><td>.62</td></tr></table>										N	Mean Score	SD	r	124	96.9	21.7	.62
N	Mean Score	SD	r														
124	96.9	21.7	.62														

Chances in 100 that a man receiving one of the above scores will achieve average or better grades in Phase I of Course E-70 (Wheeled Vehicle Automotive Mechanics), Ordnance School.

(3) Auto Mechanic Test (TC-14ar).

SCORE																	
51									91								
42									73								
27									39								
18									20								
9									8								
<table><tr><th>N</th><th>Mean Score</th><th>SD</th><th>r</th></tr><tr><td>124</td><td>31.8</td><td>11.4</td><td>.63</td></tr></table>										N	Mean Score	SD	r	124	31.8	11.4	.63
N	Mean Score	SD	r														
124	31.8	11.4	.63														

Chances in 100 that a man receiving one of the above scores will achieve average or better grades in Phase I of Course E-70 (Wheeled Vehicle Automotive Mechanics), Ordnance School.



(4) Distributor and Valves Test (TC-15a).

## SCORE

[illegible]

Chances in 100 that a man receiving one of the above scores will achieve average or better grades in Phase I of Course E-70 (Wheeled Vehicle Automotive Mechanics), Ordnance School.

(5) Use-of-Tools Test (TC-16a).

## SCORE

<b>22</b>								<b>75</b>
<b>20</b>								<b>69</b>
<b>15</b>								<b>50</b>
<b>13</b>								<b>43</b>
<b>10</b>								<b>32</b>

N	Mean Score	SD	r
124	14.9	4.6	.40

Chances in 100 that a man receiving one of the above scores will achieve average or better grades in Phase I of Course E-70 (Wheeled Vehicle Automotive Mechanics), Ordnance School.

### C. SELECTION FOR DIRECT ASSIGNMENT.

- (2) Auto Mechanic Test (TC-14ar).

127. Tank Mechanic, Minor Maintenance  
(MOS-660)

### SELECTION FOR TRAINING:

- a. Mechanical Aptitude Test.

## SCORE

[illegible]

*Chances in 100 that a man receiving one of the above scores will achieve average or better in training.*

- b. Other factors of value in selecting men for training as tank mechanics are previous occupation and education. Men with civilian experience in automotive work stand a better chance than those without. Evidence also points to the desirability of 10 years of schooling.

128. Airplane and Engine Mechanic (MOS-747)

### SELECTION FOR TRAINING:

- a. Mechanical Aptitude Test.

## SCORE

<b>140</b>										<b>87</b>
<b>120</b>										<b>68</b>
<b>100</b>										<b>42</b>
<b>80</b>										<b>19</b>
<b>60</b>										<b>6</b>

N	Mean Score	SD	r
453	106.3	15.0	.45

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

## b. Army General Classification Test.

SCORE																	
140									88								
120									67								
100									40								
80									17								
60									5								
<table><tr><th>N</th><th>Mean Score</th><th>SD</th><th>r</th></tr><tr><td>548</td><td>107.1</td><td>12.8</td><td>.41</td></tr></table>										N	Mean Score	SD	r	548	107.1	12.8	.41
N	Mean Score	SD	r														
548	107.1	12.8	.41														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

## c. Trade Information Test (TC-1a).

SCORE																	
140									83								
120									62								
100									35								
80									15								
60									5								
<table><tr><th>N</th><th>Mean Score</th><th>SD</th><th>r</th></tr><tr><td>535</td><td>111.4</td><td>18.5</td><td>.53</td></tr></table>										N	Mean Score	SD	r	535	111.4	18.5	.53
N	Mean Score	SD	r														
535	111.4	18.5	.53														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

## d. General Technical Test (TC-2a).

SCORE																	
140									88								
120									63								
100									29								
80									12								
60	3																
<table><tr><th>N</th><th>Mean Score</th><th>SD</th><th>r</th></tr><tr><td>400</td><td>113.7</td><td>15.1</td><td>.55</td></tr></table>										N	Mean Score	SD	r	400	113.7	15.1	.55
N	Mean Score	SD	r														
400	113.7	15.1	.55														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

e. Technical Trade Test (TC-7a) is composed of items selected from TC-1a and TC-2a.

## f. Nut and Bolt Manual Dexterity Test (TC-5a).

SCORE																	
140									80								
120									62								
100									39								
80									20								
60									8								
<table><tr><td>N</td><td>Mean Score</td><td>SD</td><td>r</td></tr><tr><td>504</td><td>109.5</td><td>11.2</td><td>.30</td></tr></table>										N	Mean Score	SD	r	504	109.5	11.2	.30
N	Mean Score	SD	r														
504	109.5	11.2	.30														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

## g. The U-Bolt Test (TC-6a).

SCORE																	
140									69								
120									61								
100									52								
80									42								
60									25								
<table><tr><td>N</td><td>Mean Score</td><td>SD</td><td>r</td></tr><tr><td>504</td><td>97.0</td><td>18.0</td><td>.34</td></tr></table>										N	Mean Score	SD	r	504	97.0	18.0	.34
N	Mean Score	SD	r														
504	97.0	18.0	.34														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

## 129. Airplane Armorer (MOS-911)

## SELECTION FOR TRAINING:

## a. Mechanical Aptitude Test.

SCORE																	
140									91								
120									70								
100									39								
80									14								
60									3								
<table><tr><th>N</th><th>Mean Score</th><th>SD</th><th>r</th></tr><tr><td>102</td><td>106.8</td><td>14.5</td><td>.51</td></tr></table>										N	Mean Score	SD	r	102	106.8	14.5	.51
N	Mean Score	SD	r														
102	106.8	14.5	.51														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

## b. Trade Information Test (TC-1a).

SCORE																	
140									82								
120									60								
100									35								
80									15								
60									6								
<table><tr><th>N</th><th>Mean Score</th><th>SD</th><th>r</th></tr><tr><td>130</td><td>112.2</td><td>16.8</td><td>.47</td></tr></table>										N	Mean Score	SD	r	130	112.2	16.8	.47
N	Mean Score	SD	r														
130	112.2	16.8	.47														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

## c. General Technical Test (TC-2a).

SCORE																	
140									79								
120									56								
100									30								
80									15								
60									6								
<table><tr><th>N</th><th>Mean Score</th><th>SD</th><th>r</th></tr><tr><td>128</td><td>117.2</td><td>14.8</td><td>.44</td></tr></table>										N	Mean Score	SD	r	128	117.2	14.8	.44
N	Mean Score	SD	r														
128	117.2	14.8	.44														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

## d. Technical Trade Test (TC-7a).

SCORE																	
140									80								
120									58								
100									35								
80									16								
60									5								
<table><tr><td>N</td><td>Mean Score</td><td>SD</td><td>r</td></tr><tr><td>121</td><td>113.0</td><td>16.8</td><td>.46</td></tr></table>										N	Mean Score	SD	r	121	113.0	16.8	.46
N	Mean Score	SD	r														
121	113.0	16.8	.46														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

## e. Nut and Bolt Manual Dexterity Test (TC-5a).

SCORE																	
140									76								
120									62								
100									45								
80									30								
60									17								
<table><tr><th>N</th><th>Mean Score</th><th>SD</th><th>r</th></tr><tr><td>130</td><td>105.7</td><td>12.7</td><td>.26</td></tr></table>										N	Mean Score	SD	r	130	105.7	12.7	.26
N	Mean Score	SD	r														
130	105.7	12.7	.26														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

## f. U-Bolt Assembly Test (TC-6a).

SCORE																	
140									67								
120									60								
100									51								
80									45								
60									27								
<table><tr><th>N</th><th>Mean Score</th><th>SD</th><th>r</th></tr><tr><td>130</td><td>97.0</td><td>19.1</td><td>.33</td></tr></table>										N	Mean Score	SD	r	130	97.0	19.1	.33
N	Mean Score	SD	r														
130	97.0	19.1	.33														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

## 130. Machinist (MOS-114)

## SELECTION FOR DIRECT ASSIGNMENT:

a. The Machinist Experience Check List (TC-17ar) provides a convenient method for securing a general appraisal of a man's background in machine work. The number and types of items checked by the average graduate of a machinist course are indicated in a supplement to Manual TC-M12.

b. The Machinist Test (TC-18ar) is designed to measure technical knowledge essential in machine work. Scores on the test may be used to identify soldiers in this military occupational specialty who require refresher training before reassignment. Critical scores and interpretations are given in Manual TC-M12 and supplements.



## b. Clerical Aptitude Test.

SCORE												
140				55								
120			30									
100		12										
80	4											
60	1											
<table><tr><td>N</td><td>Mean Score</td><td>SD</td><td>r</td></tr><tr><td>415</td><td>136.2</td><td>10.5</td><td>.32</td></tr></table>					N	Mean Score	SD	r	415	136.2	10.5	.32
N	Mean Score	SD	r									
415	136.2	10.5	.32									

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

## c. Cryptography Test (TC-4a).

SCORE																	
140									96								
120									82								
100									54								
80									24								
60	7																
<table><tr><th>N</th><th>Mean Score</th><th>SD</th><th>r</th></tr><tr><td>415</td><td>80.9</td><td>7.7</td><td>.61</td></tr></table>										N	Mean Score	SD	r	415	80.9	7.7	.61
N	Mean Score	SD	r														
415	80.9	7.7	.61														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

d. Other factors. In most instances, trainees and men assigned to this work must be citizens and must be approved by a G-2 investigation. A background of training or experience in mathematics or languages is generally considered desirable. In addition, the Cryptographic Technician must be able to type at the rate of 25 words per minute.

136. Radio Operator, AAF (MOS-756); Radio Operator-Mechanic, Gunner, AAF (MOS-757); Radio Operator, Low Speed, AAF (MOS-776); and Radio Operator and Mechanic, AAF (MOS-2756)

SELECTION FOR TRAINING:

## a. Mechanical Aptitude Test.

SCORE																	
140									77								
120									58								
100									36								
80									18								
60									8								
<table><tr><th>N</th><th>Mean Score</th><th>SD</th><th>r</th></tr><tr><td>698</td><td>113.0</td><td>15.7</td><td>.39</td></tr></table>										N	Mean Score	SD	r	698	113.0	15.7	.39
N	Mean Score	SD	r														
698	113.0	15.7	.39														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

## b. Army Radio Code Aptitude Test.

SCORE																	
140									74								
130									61								
120									47								
110									33								
100									21								
<table><tr><th>N</th><th>Mean Score</th><th>SD</th><th>r</th></tr><tr><td>221</td><td>122.0</td><td>18.0</td><td>.55</td></tr></table>										N	Mean Score	SD	r	221	122.0	18.0	.55
N	Mean Score	SD	r														
221	122.0	18.0	.55														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

## 137. Weather Observer (MOS-784)

SELECTION FOR TRAINING:

## a. Army General Classification Test.

SCORE																	
140									65								
120									35								
100									12								
80									3								
60									0								
<table><tr><th>N</th><th>Mean Score</th><th>SD</th><th>r</th></tr><tr><td>1042</td><td>130.1</td><td>12.5</td><td>.42</td></tr></table>										N	Mean Score	SD	r	1042	130.1	12.5	.42
N	Mean Score	SD	r														
1042	130.1	12.5	.42														

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

b. Weather Aptitude Test (TC-3a).

SCORE									
140									81
120									63
100									41
80									22
60									9

N	Mean Score	SD	r
498	108.2	12.6	.33

Chances in 100 that a man receiving one of the above scores will achieve average or better in training.

c. Other factors. Training in physics (high school or college) is essential. A knowledge of algebra and prior training and/or experience in meteorology are valuable assets.

**138. Officer Candidate (MOS-625)**

### SELECTION FOR TRAINING:

a. Army General Classification Test.

SCORE					
140					72
120					56
100				39	
80	*				
60	*				

N	Mean Score	SD	r
306	126.5	10.0	.38

Chances in 100 that a man receiving one of the above scores will achieve average or better academic grades in training.

\*A score of 110 on AGCT is required for selection as a candidate for officer training.

b. Officer Candidate Test (OCT-1 and OCT-2).

SCORE		Frequency			
140	83				
120	62				
100	37				
80	17				
60	5				

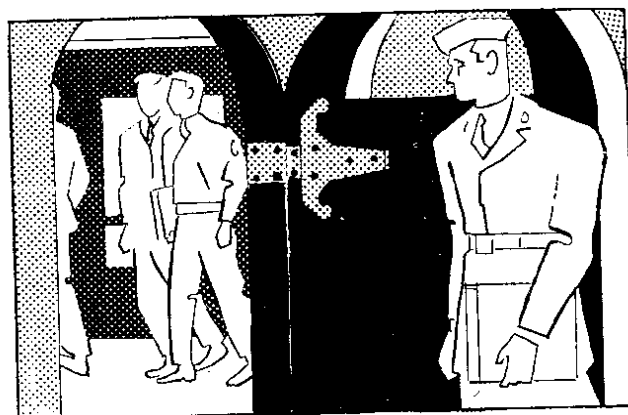
  

N	Mean Score	SD	r
306	125.2	12.4	.62

Chances in 100 that a man receiving one of the above scores will achieve average or better academic grades in training.

c. Other factors. Next to academic proficiency, leadership is of paramount importance for candidates for officer training. Weakness in one or the other of these two areas accounts for 75 percent—90 percent of all failures in Officer Candidate Schools. Until the qualities of leadership have been isolated and more objectively defined, no tests can be developed for estimating these qualities. Nevertheless, evaluations of leadership traits by OCS selection agencies can be rendered more standard and objective by a proper utilization of *WD AGO Form 240 (Interviewing and Rating of Applicants for Officer Candidate School)*. The principles of interviewing and rating outlined in Chapter 6 of this manual are applicable in this connection.

## Section V. OTHER TESTS



### 139. Individual Test of General Mental Ability

a. The Army Individual Test is designed for use in connection with personnel problems in which it appears likely that an individual instrument will provide a better estimate of general learning ability than the usual group test. It measures the same general abilities as the Army General Classification Test. In addition, since it is administered individually and involves a variety of verbal and performance materials, it provides the examiner an opportunity to control motivation more effectively and to observe and evaluate the examinee's behavior more fully.

b. The Individual Target Test is a means of determining aptitude for assimilating basic military training among men who are slow learners, illiterate or semi-illiterate or limited in verbal and numerical skills. Scores are interpreted in terms of probable soldier proficiency of the examinees. It is applicable only to men with limited general mental ability, and should be employed, in such cases, as an objective adjunct to other available means of estimating chances for success in basic military assignments.

c. The Nonlanguage Test, 2abc is designed to grade Army personnel in terms of their ability to

learn Army duties. As the name implies, the test minimizes the use of language. It is composed of three subtests involving box counting, symbol association and substitution, and design comparison. It is administered as a group test, all directions being given in pantomime. Scores on the test are distributed in much the same fashion as AGCT scores, and the relationship between the two tests is fairly close. The Nonlanguage Test, 2abc was formerly given in reception centers to all men who scored Grade V on the AGCT. Since such men are now assigned to Special Training Units, its use is now largely confined to training centers and units where it is occasionally employed as a check on low AGCT scores.

### 140. Warrant Officer Examinations

Applicants for appointment as warrant officers must pass written technical examinations in addition to satisfying the other requirements of experience and training. Separate examinations are available for each of the various classifications in which warrant officers may be appointed. The duties involved in each classification and the scope of the appropriate examinations are outlined in AR 610-10 and 610-15.

## INDEX

	Paragraph	Page		Paragraph	Page
Ability, an equation.....	5	4	Cook experience check list (TC-27ar).....	134	84
Achievement tests.....	22, 117	13, 76	Cook test (TC-28ar).....	134	84
Administering of tests.....	29	17	Criteria—		
Aircraft observer, ground (MOS-518).....	133	84	Of success.....	66	46
Aircraft warning aptitude test (TC-10a).....	133	84	Relation to actual performance.....	32	19
Aircraft warning classification test (TC-11a).....	133	84	Validity.....	32	19
Aircraft warning specialists.....	133	84	Critical scores.....	37, 69, 73	23, 48, 50
Airplane and engine mechanic (MOS-747).....	128	81	Cryptanalysis technician (MOS-808).....	135	84
Airplane armorer (MOS-911).....	129	82	Cryptographic code compiler (MOS-807).....	135	84
Aptitude:			Cryptographic technician (MOS-805).....	135	84
Individual differences.....	9	5	Cryptography test (TC-4a).....	135	84
Tests.....	22, 116	13, 76	Dictation tests.....	123	79
Army automotive screening test battery.....	126	80	Discriminative index (items).....	32	19
Army general classification—			Distributor and valves test (TC-15a).....	128	80
Test-1.....	113	74	Duties of proctors.....	45	28
Test-3.....	12, 111	10, 72	Expectancy charts.....	68, 73, 119	47, 50, 77
Army grades.....	63	43	And class average.....	119	77
Army individual test.....	139	87	And course content.....	119	77
Army radio code aptitude test.....	22, 112, 136	13, 73, 85	And course standards.....	119	77
Army standard scores—			Field trials, of tests.....	32	19
Advantages.....	62	43	Final form of tests.....	34	22
Are not "IQ's".....	64	45	General technical test (TC-2a).....	128	81
Conversion from raw.....	58	39	Group target test (GT-1).....	102	62
Army standard score scale:			Group testing session:		
Definition.....	60, 61	41	Administration.....	45	28
Development.....	61	41	Collection and disposition of test ma-		
Army tests:			terials.....	45	28
As scientific tools.....	14	11	Preparation.....	45	28
Ease of administration.....	19	12	Group tests.....	23	14
Economy of time basic to.....	19	12	"Halo" effect in ratings.....	82	54
Reliability.....	17	11	Individual examination (IE-1).....	102	62
Significance.....	18	11	Individual differences, physical.....	9	5
Uniformity.....	16	11	Individual target test.....	139	87
Assignments:			Individual testing session:		
Relation to training.....	9	5	Administration.....	52	36
Types.....	8	5	Qualifications of examiner.....	52	36
Auto mechanic test (TC-14ar).....	126	80	Timing in.....	54	37
Automotive mechanic experience, check			Individual tests.....	23	14
list (TC-13a).....	126	80	Induction station:		
Block counting test.....	106	66	Preliminary screening interview.....	101	62
Carpenter (MOS-050).....	131	84	Psychological standards.....	100	62
Carpenter experience check list (TC-25ar).....	131	84	Standards.....	98	62
Carpenter test (TC-26ar).....	131	84	Terminal interview.....	107	68
Check List (see also Questionnaire), value.....	77	51	Testing problem.....	102	62
Classification:			Test procedures (summary).....	104	66
Definition.....	8	5	Tests.....	102	62
Measuring instruments used.....	14	11	Previously authorized.....	106	66
Purpose.....	2	1	Recording scores.....	103	65
Relation to tests.....	20	13	Validity.....	105	66
Clerical achievement test (TC-24ar).....	121	78	Information center operator (MOS-510).....	133	84
Clerical aptitude test.....	121	78	Instructions, importance of test.....	41	25
Clerical experience check list (TC-23ar).....	121	78	International test scoring machine.....	51	35
Clerk, general (MOS-055).....	121	78			
Clerk-Typist (MOS-405).....	122	79			
Conversion tables.....	61d, 73	43, 50			
Cook (MOS-060).....	134	84			



## INDEX

	Paragraph	Page		Paragraph	Page
Interview:			Qualification test (Q-1 and Q-2)	102	62
Comparison with other measures	79	53	Questionnaire:		
Conduct	96	57	Comparison with other measures	79	53
Definition	91	59	Construction	88	57
Preliminary screening	101	62	Definition	85	56
Purpose	92	60	Information obtained from	86	57
Reliability	93	60	Kinds	87	57
Terminal	107	68	Preliminary trial	89	59
Value	78,94	51,60	Scoring	90	59
Interviewers:			Value	77	51
Preparation	96	61	Radar crewman (MOS-514)	133	84
Qualification of Army	95	61	Radio operator, AAF (MOS-756)	136	85
Interviewing and rating of applicants for			Radio operator, low speed (MOS-776)	136	85
OCS (WD AGO Form 240)	138	86	Radio operator and mechanic, AAF		
Items (test):			(MOS-2756)	136	85
Constructing	28	17	Radio operator-mechanic, gunner, AAF		
Difficulty	32	19	(MOS-757)	136	85
Discriminative index	32	19	Radiotelegraph operator aptitude test		
Free answer	24	15	(ROA-1, X-1)	113	74
Multiple choice	24	15	Rating scale:		
Selection	33	21	Administration	84	56
Validity	33	21	Comparison with other measures	79	53
Length of tests	25	16	Construction	82	54
Machinist (MOS-114)	130	83	Defining position on	82	54
Machinist experience check list (TC-17ar)	130	83	Definition	80	53
Machinist test (TC-18ar)	130	83	Degrees in	82	54
Measurement	6	4	Graphic	81	53
Characteristics of scientific	13	10	"Halo" effect in	82	54
Psychological	11,12	9,10	Numerical	81	53
Mechanical aptitude test	113,126	74,80	Reliability	83	55
Mechanic, automotive (second echelon)			Scoring	84	56
(MOS-014)	126	80	Types	81	53
Mechanic, automotive, track vehicle			Validity	83	55
(third echelon) (MOS-966)	126	80	Value	76	51
Mechanic, automotive, wheel vehicle			Reception center:		
(third echelon) (MOS-965)	126	80	Classification interview	109	68
Nonlanguage individual examination			Testing program	110	72
(NIE-1)	102	62	Reliability:		
Nonlanguage test, 2abc	139	87	Coefficient	72,73	49,50
Nut and bolt manual dexterity test			Definition	70	49
(TC-5a)	128	81	Estimating test	71	49
Objectives of tests	21	13	Improving test	72	49
Officer candidate (MOS-625)	138	86	Of Army tests	17	11
Officer candidate test (OCT-1 and OCT-2)	138	86	Retesting	72	49
Oral trade questions (uses)	109	68	Scientific method, need	4	1
Performance tests	23	14	Scores:		
Power tests	26	17	And prediction of performance	66	46
Proctors, duties	45	28	Distribution	61	41
Psychological measurement:			Interpretation	72	49
Indirect observation	12	10	Percentile	63	43
Methods	11	9	Range	61	41
Facial analysis	11	9	Standard	58,62	39,43
Graphology	11	9	Scoring group tests:		
Phrenology	11	9	Hand	47,49	31,33
			Of separate answer sheets	50	33

## INDEX

	Paragraph	Page		Paragraph	Page
Scoring group tests—Continued:			Tests—Continued:		
Machine.....	47	31	Field trials.....	32	19
Of answer sheets.....	51	35	Final form.....	34	22
Preparation of answer sheets.....	51	35	Group.....	23	14
Scoring individual tests:			Individual.....	23	14
Criteria.....	57	38	Induction station.....	102	62
Special answer sheet.....	56	38	Length.....	25	16
Scoring tests:			Objectives.....	21	13
Directions.....	30	18	Performance.....	23	14
Formula.....	30	18	Populations.....	32	19
Selection:			Power.....	26	17
Measurement and.....	9	5	Reception center.....	110	72
Value of tests for.....	118	76	Reliability.....	71, 72	49
Skills:			Scoring formula.....	30	18
Individual differences.....	9	5	Speed.....	26	17
Level reached with training.....	9	5	Standard.....	58	39
Measurement.....	15	11	Standardization.....	35	22
Soldier's Qualification card (WD AGO			Standardization population.....	36	23
Form 20).....	109	68	Time limits.....	26	17
Speed tests.....	26	17	Timing.....	45	28
Standardization:			Validity.....	66, 67	46, 47
Definition.....	42	25	Verbal.....	23	14
Of tests.....	35	22	Tool usage film strip test (TC-12a).....	126	80
Population.....	36	23	Trade information test (TC-1a).....	128	81
Standards:			Truck driver experience check list (TC-		
For induction.....	98	62	21ar).....	125	79
Psychological.....	100	62	Truck driver, heavy (MOS-931).....	125	79
Standard scores, conversion from raw to.....	58	39	Truck driver, light (MOS-345).....	125	79
Stenographer (MOS-213).....	123	79	Truck driver test (TC-22ar).....	125	79
Supply clerk (quartermaster supplies)			Typing tests.....	122	79
(MOS-835).....	124	79	U-bolt test (TC-6a).....	128	81
Supply clerk experience check list (quarter-			Use-of-tools test (TC-16a).....	126	80
master supplies) (TC-29ar).....	124	79	Validity:		
Supply clerk test, (quartermaster sup-			Coefficient.....	67, 73	47, 50
plies) (TC-30ar).....	124	79	Definition.....	66	46
Tank mechanic, minor maintenance			Verbal tests.....	23	14
(MOS-660).....	127	81	Visual classification test.....	106	66
Technical trade test (TC-7a).....	128	81	Warrant officer examinations.....	140	87
Testing situation, physical surroundings.....	44	27	WD AGO Form 20 (Soldier's Qualifi-		
Tests:			cation Card).....	109	68
Achievement.....	22	13	WD AGO Form 240 (Interviewing and		
Administration.....	29	17	Rating of Applicants for OCS).....	138	86
Aptitude.....	22	13	Weather aptitude test (TC-3a).....	137	85
Classification.....	20	13	Weather observer (MOS-784).....	137	85
Comparison with other measures.....	79	63	Welder, combination (MOS-256).....	132	84
Construction.....	29	17	Welding experience check list (TC-19ar).....	132	84
Continuing studies.....	38	23	Welding test (TC-20ar).....	132	84
Critical scores on.....	37	23	Wells' concrete directions test.....	106	66
Directions for scoring.....	30	18			